Lingual vs. labial fixed orthodontic appliances: systematic review and meta-analysis of treatment effects

Papageorgiou, Spyridon N ; Gölz, Lina ; Jäger, Andreas ; Eliades, Theodore ; Bourauel, Christoph

Abstract: The aim of this systematic review was to compare the therapeutic and adverse effects of lingual and labial orthodontic fixed appliances from clinical trials on human patients in an evidence-based manner. Randomized and prospective non-randomized clinical trials comparing lingual and labial appliances were included. Risk of bias within and across studies was assessed using the Cochrane tool and the Grading of Recommendations Assessment, Development and Evaluation (GRADE) approach. Random-effects meta-analyses were conducted, followed by subgroup and sensitivity analyses. Six electronic databases were searched from inception to July 2015, without limitations. A total of 13 papers pertaining to 11 clinical trials were included with a total of 407 (34% male/66% female) patients. Compared with labial appliances, lingual appliances were associated with increased overall oral discomfort, increased speech impediment (measured using auditory analysis), worse speech performance assessed by laypersons, increased eating difficulty, and decreased intermolar width. On the other hand, lingual appliances were associated with increased intercanine width and significantly decreased anchorage loss of the maxillary first molar during space closure. Based on existing trials, there is insufficient evidence to make robust recommendations for lingual fixed orthodontic appliances regarding their therapeutic or adverse effects, as the quality of evidence was low.

DOI: https://doi.org/10.1111/eos.12250

Title Page

Lingual versus labial fixed orthodontic appliances: systematic review and meta-analysis of treatment effects

Spyridon N. Papageorgiou[1,2], Lina Gölz[1], Andreas Jäger[1], Theodore Eliades[3], Christoph Bourauel[2]

[1]Department of Orthodontics, School of Dentistry, University of Bonn, Welschnonnenstr. 17, 53111, Bonn, Germany; [2]Department of Oral Technology, School of Dentistry, University of Bonn, Welschnonnenstr. 17, 53111, Bonn, Germany; [3]Clinic of Orthodontics and Paediatric Dentistry, Center of Dental Medicine, Faculty of Medicine, University of Zurich, Plattenstrasse 11, Zurich 8032, Switzerland

Running title: Lingual versus labial fixed appliances

Corresponding author: Spyridon N. Papageorgiou, Department of Orthodontics, School of Dentistry, University of Bonn, Welschnonnenstr. 17, 53111, Bonn, Germany. Tel.: +49-(0)228-287-22449; Fax: +49-(0) 228-287-22588; E-mail: snpapage@gmail.com.

Words in abstract: 200
Words in text: 4064

Keywords: orthodontics, fixed appliances, lingual appliances, labial appliances, systematic review, meta-analysis

Papageorgiou SN, Gölz L, Jäger A, Eliades T, Bourauel C. Lingual versus labial fixed orthodontic appliances: systematic review and meta-analysis of treatment effects. Eur J Oral Sci 2015.

Abstract

Aim of this systematic review was to compare the therapeutic and adverse effects of lingual and labial orthodontic fixed appliances from clinical trials on human patients in an evidence-based manner. Randomized and prospective non-randomized clinical trials comparing lingual and labial appliances were included. Risk of bias within and across studies was assessed with the Cochrane tool and the GRADE approach. Random-effects meta-analyses were conducted, followed by subgroup and sensitivity analyses. Six electronic databases were searched from inception to July 2015 without limitations. A total of 13 papers pertaining to 11 clinical trials were included with a total of 407 (119 male / 228 female) patients. Lingual appliances were associated with increased overall oral discomfort compared to labial appliances, increased speech impediment (measured with auditory analysis), worse speech performance assessed by laypersons, increased eating difficulty, and decreased intermolar width. On the other side, lingual appliances were associated with increased intercanine width and significantly decreased anchorage loss of the maxillary first molar during space closure. Based on existing trials, there is insufficient evidence to make robust recommendations for lingual fixed orthodontic appliances regarding their therapeutic or adverse effects, as the quality of evidence was low.

Registration: PROSPERO (CRD42015024596)

Conflict of interest: None

Manuscript

Introduction

Rationale

Fixed appliance treatment has become an integral part in modern orthodontics and has been a major focus point of orthodontic research. Traditionally, orthodontic appliances have been fixed on the outer (labial) surface of the teeth (hereon labial appliances). In recent years, the increased number of adult patients seeking orthodontic treatment (1) and their higher esthetic demands (2) have led to the development of various esthetic treatment approaches including esthetic brackets, clear aligners, and appliances fixed on the inner (lingual or palatal) surface of the teeth (hereon lingual appliances).

Since introduction of lingual appliances by Fujita (3), progress has been seen in their design, manufacturing, and mechanotherapy. Advantages of lingual appliances proposed by clinicians or manufacturers include lower noticeability, fewer white spot lesions and caries, lighter forces being needed due to smaller interbracket distance, smaller anchorage loss, and increased comfort (2, 4, 5). Possible disadvantages include practical difficulties in the insertion and handling of these appliances, longer chairtimes for patients and orthodontists, higher laboratory costs, and poorer outcomes compared to labial appliances. The development of new archwire materials, advanced laboratory techniques, and the widespread use of sophisticated computer programs have reintroduced lingual appliances as a promising and a competing technique by trying to alleviate or overcome some of the abovementioned disadvantages.

Existing systematic assessment of orthodontic fixed appliances are limited and problematic (6, 7). Current evidence on lingual appliances has been previously quantitatively assessed (8, 9). However, conclusions may have been distorted by inclusion of retrospective studies (10), limited identification of eligible trials, or issues during their qualitative/quantitative data synthesis (11, 12). In particular, assessment of the quality of evidence with the Grades of Recommendation, Assessment, Development and Evaluation (GRADE) approach (13) and their translation in future clinical settings (14) could aid in drawing robust conclusions.

Objectives

Aim of this study was to compare the treatment effects of lingual appliances compared to labial appliances from randomized and prospective non-randomized clinical trials conducted on human patients.

Materials and Methods

Protocol and registration

The protocol for this review was made *a priori* based on the PRISMA-P statement (15), registered in PROSPERO (CRD42015024596), and all *post hoc* changes were appropriately noted. This systematic review was conducted and reported according to Cochrane Handbook (16) and PRISMA statement (17), respectively.

Eligibility criteria

According to the PICOS schema, included were parallel or split-mouth randomized and non-randomized prospective controlled trials on human patients comparing any lingual appliance to any labial appliance and assessing its therapeutic effects (both effectiveness and efficiency) or adverse effects. Excluded were non-clinical studies, retrospective studies, and studies with partial appliances (appliance not placed on all teeth excluding second and third molars).

Information sources and literature search

A total of six electronic databases were searched systematically by one author (SNP) without any limitations from inception up to July 20th, 2015 (Appendix 1). Four additional sources (Scopus, Google Scholar, ClinicalTrials.gov, and ISRCTN registry) were manually searched for additional trials or protocols by the same author. Authors of included trials were contacted for additional missed or ongoing trials. No limitations concerning language, publication year or status were applied. The reference lists of the included trials and relevant reviews were manually searched as well.

Study selection

Titles identified from the search were screened by one author (SNP) with a subsequent duplicate independent checking of their abstracts/full-texts against the eligibility criteria by two authors (SNP, LG), while conflicts were resolved by a third author (AJ).

Data collection

Characteristics of included trials and numerical data were extracted in duplicate by two authors (SNP, LG) using pre-determined and piloted extraction forms. Piloting of the forms was performed during the protocol stage until over 90% agreement was reached. Missing or unclear information was requested by the trials' authors.

Risk of bias in individual trials

The risk of bias of the included trials was assessed using Cochrane's risk of bias tool (16) after initial calibration. A main risk of bias assessment was included in the systematic review pertaining to each trial's primary outcome.

Data synthesis

As the outcome of fixed appliance therapy is bound to be affected by the bracket (7), the wire (6), and their interaction (18), a random-effects model according to DerSimonian and Laird was deemed appropriate to incorporate this variability (11).

For parallel trials the Mean Difference (MD) and the Relative Risk (RR) with their corresponding 95% Confidence Interval (CI) were chosen as effect measures for continuous and binary outcomes, respectively. The RR was chosen over the Odds Ratio (OR), due to its comparative advantages (18). For split-mouth trials the raw data were requested from the trial's authors and clustering-adjusted estimates were calculated with univariable and multivariable regression modeling. In case similar outcomes were assessed both as binary and continuous, the Standardized Mean Difference (SMD) was chosen to pool them after conversion according to Chinn (20). The number needed to treat was planned to be used to clinically translate the results of statistically significant meta-analyses of binary outcomes.

Between-trial heterogeneity was quantified with the I^2 statistic, defined as the proportion of total variability in the results explained by heterogeneity, and not chance (21, 22). The 95% uncertainty intervals (95% UI) (similar to CIs) around the I^2 were calculated (23) using the non-central χ^2 approximation of Q (24). 95% predictive intervals were calculated for meta-analyses of three trials or more, which incorporate existing heterogeneity and provide a range of possible effects for a future clinical setting (14). All analyses were run in Stata SE 10.0 (StataCorp, College Station, TX) by one author (SNP). A two-tailed P-value of 0.05 was considered significant for hypothesis-testing, except for a 0.10 used for the test of heterogeneity and reporting biases, due to low power (25).

Risk of bias across studies

The overall quality of evidence (confidence in effect estimates) for each of the main outcomes was rated using the GRADE approach (13). For this assessment, the risk of bias of each included trial was re-assessed separately at outcome level.

The minimal clinical important, large, and very large effects were conventionally defined (26) as half, one, and two standard deviations, respectively. The standard deviation for an outcome was averaged from the existing trials. Conventional cut-offs of 0.2, 0.5, and 0.8 were adopted for the SMD. The produced forest plots were augmented with contours denoting the magnitude of the observed effects. Finally, the optimal information size (i.e. required meta-analysis sample size) was calculated for each outcome independently for $\alpha = 5\%$ and $\beta = 20\%$.

Additional analyses

Possible sources of heterogeneity were planned to be sought through pre-specified mixed-effects subgroup analyses and random-effects meta-regression with the Knapp-Hartung adjustment (27), should at least five trials be pooled. Indications of reporting biases (including small-study effects) were planned to be assessed with Egger's linear regression test (28) and contour-enhanced funnel plots, should ten or more trials be pooled.

Sensitivity analyses

As prospective non-randomized trials were also planned to be included in addition to randomized controlled trials (RCTs), a sensitivity analysis was planned to be conducted by including only RCTs and compared with the original analysis. Additionally, a *post hoc* exploratory analysis was performed to assess the overall difference between randomized and non-randomized trials on lingual appliances, adopting a quantitative approach (10). All meta-analyses were converted to SMDs on the same effect direction and differences according to the trial design were expressed as differences in SMDs (ΔSMDs) through random-effects meta-regression and pooled across meta-analysis via random-effects meta-analysis. Additional sensitivity analyses were planned, but were not conducted due to the limited number of included trials.

Results

Study selection

A total of 123 and 7 papers were identified through the electronic (Appendix 1) and manual searches, respectively (Fig. S1). After removal of duplicates and initial screening, 31 papers were judged against the eligibility criteria, leaving a final number of 13 included papers (29, 41), (Fig. S1; Appendix 2; Table 1). In two instances duplicate publications pertaining to the same trial were grouped together, leaving a total of 11 finally included trials.

Study characteristics

The characteristics of the included trials can be seen in Table 1 and Appendix 3. Out of the 11 included trials, three (27%) were parallel RCTs, one (9%) was split-mouth RCT according to treated jaw, and the remaining seven (64%) were parallel prospective non-RCTs. They included a total of 407 patients (with at least 119 male and 228 female patients) with an average age of 21.3 years. The majority (73%) of the lingual appliance groups and all (100%) of the labial appliance groups pertained to pre-fabricated appliances, while three trials (23%) used individualized lingual appliances for each patient (Incognito® appliance, 3M-Unitek, Monrovia, Calif; formerly, TOP Service for Lingualtechnik, Bad Essen, Germany). Only six trials (55%) reported, even partially, information on the used archwires. Out of these six trials, four of them (66%) used pre-fabricated archwires for both lingual and labial groups, while two trials (33%) used individualized archwires for the lingual and pre-fabricated archwires for the labial group.

Risk of bias within studies

The risk of bias assessment for the eleven included trials can be seen in Fig. 1 and Appendix 4. Serious methodological inadequacies were found in all trials for at least one bias domain. Most problematic domains were the inadequate/inexistent randomization (high risk in 64% of the trials) and blinding of outcome assessors (missing in 73% of the trials).

Results of individual studies and data synthesis

The results of all individual included studies are quantitatively represented in Appendix 5, while the results of all performed meta-analyses with 2 or more studies are given in Table 2. In all instances the MD and the RR was used for continuous and binary outcomes, respectively. In one instance crude and adjusted ORs and incidence rate ratios were used to express the raw trial data of van der Veen et al. (38) that were re-analyzed with univariable and multivariable binomial/negative binomial regression modeling. For the outcome of oral

discomfort, the binary questionnaire measurements of Khattab et al. (30) and Caniklioglu and Oztürk (29) were combined with the continuous Likert scale measurements of Shalish et al. (35) by expressing all three trials in SMDs.

Shortly, lingual appliances were associated with the following beneficial effects ($P < 0.05$): less appliance noticeability, less discomfort at the cheeks, greater increase of the intercanine width with subsequent less needed interproximal enamel reduction, less anchorage loss of the posterior segment during space closure, and less white spot lesions compared to labial appliances. On the other side, lingual appliances were associated with the following detrimental effects ($P < 0.05$): greater oral hygiene problems (food impaction), worse oral hygiene (greater plaque index), greater tongue discomfort, greater oral pain, greater oral discomfort, greater overall irritation of the soft tissues, greater general activity problems, greater sleep disturbance, worse speech performance (measured as a greater impact on the upper frequency of the /s/ sound via auditory analysis, as well as assessment by specialists or laypersons), greater perception of articulation change, greater avoidance of certain types of conversations, and greater eating problems. However, the vast majority of comparisons were informed from a single included trial.

Risk of bias across studies

This paper is mainly focused on seven primary outcomes that were selected for assessment in the GRADE analysis (Table 3; Appendix 6): patient-reported oral discomfort, objective speech performance by measuring the upper frequency of the /s/ sound via auditory analysis, subjective speech performance assessed by laypersons, eating difficulty, intercanine width, intermolar width, and sagittal anchorage loss.

Lingual appliances were associated with increased overall oral discomfort compared to labial appliances, greatly increased speech impediment (measured with auditory analysis or assessed from laypersons), greatly increased eating difficulty, significantly increased intercanine width, slightly decreased intermolar width, and significantly decreased sagittal anchorage loss of the maxillary first molar during space closure (Fig. 2). However, the quality of all included analyses was judged very low, due to high risk of bias of included trials, inconsistency, and imprecision.

Additional analyses

Subgroup analyses and assessments of reporting biases were planned, but could not be performed due to the limited number of trials included in the meta-analyses.

The sensitivity analysis according to the improvement of the GRADE score by including only randomized trials is seen in Appendix 7. Apart from two outcomes that no randomized trials were eligible, the sensitivity analysis indicated that non-randomized trials considerably underestimated the difference between lingual and labial brackets. Indeed, the results in two out of five of the remaining outcomes were not statistically significant in the original analysis, but were statistically significant in the sensitivity analysis. Finally, the overall comparison of randomized and non-randomized trials indicates that the latter report significantly more beneficial effects of lingual appliances compared to the former (Appendix 8). This can be regarded as evidence of bias stemming from prospective non-randomized trials, the magnitude of which is considered as very large (ΔSMD = -1.28; 95%CI = -2.24,-0.32; P = 0.009).

Discussion

Summary of evidence

This systematic review included 4 randomized and 7 non-randomized trials and a total of 407 patients. A considerable lack of evidence exists regarding the therapeutic effects of lingual appliances, especially pertaining to the long-term. Most trials are small non-randomized trials that investigate short-term adverse effects with serious limitations in their planning, conduct, and reporting.

Lingual appliances were associated with higher overall oral discomfort compared to labial appliances. However, caution is indicated in the interpretation of this finding as the original and the sensitivity analysis agreed on the direction, but not on the magnitude of this effect, resulting in a GRADE of very low to moderate (Appendix 7). Additionally, the localization of discomfort was different between lingual and labial appliances. Specifically, patients with lingual appliances were 58% less likely to report discomfort on the cheeks and 238% more likely to report discomfort on the tongue compared to patients with labial appliances (Appendix 5), which agrees with previous reports (42, 43). Finally, the overall pain intensity reported at the first two treatment weeks was significantly higher in patients with lingual appliances compared to patients with labial appliances (MD = 11.9 mm in Visual Analogue Scale; Appendix 5). It should be taken in mind that oral discomfort and pain experience during orthodontic treatment is associated with patient age, personal values, and expectations (44, 45), while the majority of oral discomfort occurs within the first month and diminishes afterwards (46). Some authors have suggested that low-profile brackets should restrict less the functional space of the tongue and induce less discomfort. This was not however the case, as the results of Shalish et al. (35) that used low-profile

Incognito® brackets did not differ from the results of Caniklioglu and Oztürk (29) that used bulkier Ormco 7th Generation® brackets (SMDs of 0.61 and 0.62, respectively; *Fig.* 2).

Lingual appliances were associated with an increased speech impediment compared to labial appliances, which was seen both via auditory analysis of the /s/ sound and via subjective judgment of speech pathologists or laypersons on a Likert scale (Appendix 5). Again however, caution is indicated, as the GRADE for these meta-analyses was very low to moderate and effects might have been underestimated (Appendix 7). Additionally, patients with lingual appliances were more likely to report a perception of articulation change and avoidance of some types of conversations after 3 months compared to patients with labial appliances (Appendix 5). This agrees with previous reports (42, 43), although speech disturbances might depend on the language used in each trial (29). However, the /s/ sound that was chosen for this analysis is especially sensitive and common in most languages (47), indicating robustness of this method. The pathomechanism of speech impairment during lingual appliance therapy stems from the contact area of the tongue being shifted further palatally due to the existence of lingual brackets (48). Speech disturbances induced by lingual brackets might be associated with the brackets' design (30) and it is worth noting that speech disturbances may lead to greater social embarrassment than visible labial brackets (41).

Patients with lingual appliances were considerably more likely to report eating difficulties compared to patients with labial appliances (435% to 800% more likely, according to the original and the sensitivity analysis, respectively). Again, eating difficulties are mainly reported in the majority of patients within the first month mainly (46). A possible explanation for prolonged eating difficulties might be the posterior disocclusion caused by the bite planes incorporated on the maxillary anterior lingual (palatal) brackets that were used (29).

Treatment with lingual appliances was associated with a distinct increase in the intercanine width and decrease in the intermolar width of treated patients (Appendix 5). However, these results must be seen with caution. First of all, the GRADE quality was very low to low, mainly due to small samples and inconsistency. Additionally, dental arch dimensions are not directly relevant to the type of fixed appliances, as the influence of specific treatment mechanics and of the archwire properties (7), might act as a confounder. Indeed, both included trials were problematic in this aspect. Khattab et al. (31) used pre-fabricated wires for the labial group, but individualized wires for the lingual group. On the other hand, Soldanova et al. (37) used wire sequences that differed in the material, size, and cross-section between the lingual and labial groups, which can affect the results (7) and introduce bias. Regarding the increased intercanine width, the prominence premolar offset incorporated in the lingual wire together with the small interbracket distance in the anterior region might be the

explanation (31). Regarding the decrease in intermolar distance, a possible explanation might be lingual appliances causing irritation of the tongue, moving it to a more posterior and inferior position, and thereby affecting the force equilibrium at the posterior teeth (31). For this reason, measures to increase the transverse molar anchorage with lingual appliances have been suggested by some authors (31).

Lingual appliances were associated with significantly less sagittal anchorage loss (mesial movement) of the first maxillary molar after en masse retraction to close first premolar extraction spaces compared to labial patients (MD = -0.82 mm). Possible explanations for this pertain to smaller arch perimeter with lingual appliances leading to higher wire rigidity and better anchorage control during retraction (49), increased anchorage value of the posterior teeth due to nearness of the lingual brackets to the center of tooth resistance (5), and the force direction during space closure with lingual appliances, which leads to cortical bone anchorage due to buccal root torque and distal rotation of the molar crown (5). Although this effect was non-trivial, confirmatory RCTs are needed.

Although not included in the GRADE analyses, due to space limitations, several other possibly significant differences were found between lingual and labial appliances. Lingual appliances were associated with a minimal, but distinct, worsening of oral hygiene (higher plaque index) compared to labial appliances (Appendix 7). Plaque deposits on the lingual gingival margins of the teeth might be more difficult to remove, especially with wider brackets and reduced interbracket distance (48), and if maintained, can cause gingival inflammation. These findings can be supported by previous studies reporting oral hygiene impairment (29, 50) and elevated plaque accumulation and gingivitis in patients with lingual appliances (42, 43).

The impact of fixed appliances on the formation of new white spot lesions during orthodontic treatment was directly re-calculated from the raw data of van der Veen et al. (38), by taking into account within-patient correlations stemming from the split-mouth design of the trial. After adjusting for all possible confounders through multivariable regression, patients' jaws treated with lingual appliances were associated with significantly fewer new white spot lesions compared to jaws treated with labial appliances (incidence of new white spots per patient jaw decreased by 72%; Appendix 5). Caries is a multifactorial phenomenon and its occurrence during orthodontic treatment is affected by the lower resting pH, increased volume of plaque, and an appliance-induced rapid shift in bacterial flora (32, 51). An explanation usually suggested for the lower caries risk with lingual appliances is the mechanical cleaning of the tongue on the lingual/palatal surfaces of the teeth, although the opposite was found from the present review. Another more viable explanation is an increased saliva flow on the lingual/palatal tooth surfaces keeping the pH high (52). In the single trial assessing this (32),

lingual appliances were associated with decreased salivary flow rate and buffering capacity, although this was not statistically significant (Appendix 5). Additionally, lingual appliances were associated with increased counts of *Streptococcus mutans* and *Lactobacilli*, but this was also not statistically significant, due to low sample size (Appendix 5). Caries activity has been negatively associated with increased counts of *Streptococcus mutans* and *Lactobacilli* (53, 54). It is however known that orthodontic treatment is accompanied by a transient short-term elevation of *Streptococcus mutans* levels, which decrease after the active treatment phase, and return to physiological levels after the removal of the retention appliances (55).

Finally, treatment with lingual appliances was associated with smaller amounts of interproximal enamel reduction needed to create missing arch space compared to labial appliances (MD = -0.67 mm; Appendix 5). However, this outcome is directly associated with initial crowding, treatment-induced changes on the dental arch width (and subsequently the arch's circumference), and the treatment protocol. As these factors were not taken into account in the original analysis, residual bias for this outcome cannot be ruled out.

Strengths and limitations

The strengths of this systematic review include the extensive unrestrictive literature search, the robust review procedures, the communication with trialists for clarifications, and the attempt to acquire and re-analyze appropriately the raw data of the included trials, as was done in the split-mouth trial of van der Veen et al. (38). Finally, this review improves on previous similar studies, as it was registered *a priori*, compared directly lingual to labial appliances, did not include biased retrospective trials, provided quantitative data for all included studies, assesss the quality of evidence with the GRADE approach, and the sensitivity analysis identified detrimental factors for the quality of clinical recommendations.

However, there exist also some limitations to this study. First and foremost, this systematic review could potentially suffer from the GIGO (garbage-in-garbage-out) principle. This pertains to the fact that the quality of existing trials comparing lingual and labial appliances is problematic, while mainly non-randomized trials exist. This might potentially influence the magnitude and direction of observed effects (10, 56, 57), as was seen firsthand in the performed sensitivity analyses. Furthermore, additional outcome data from trialists could not be obtained, apart from one instance. Moreover, the assessment of lingual and labial appliances could not be assessed in conjunction with (a) patient gender, (b) patient motivation, (c) whether the appliances were fully individualized or not, and (d) whether a direct or indirect bonding protocol was followed, although originally

planned. Finally, the limited number of included trials precluded robust assessments of heterogeneity, subgroup analyses, small-study effects and reporting biases.

There is insufficient evidence at present to make robust recommendations for lingual fixed orthodontic fixed appliances regarding their therapeutic or adverse effects. Only two out of the eleven identified trials were randomized, while none were in low risk of bias.

Recommendations for clinical practice

Due to the fact that the confidence in effect estimates from the original analysis is so low (very low GRADE), making recommendations based on them might be too speculative. Based on the effect estimates from the sensitivity analysis (moderate GRADE), orthodontists might reasonably expect more oral discomfort, speech impairment, and eating difficulty in patients with lingual appliances compared to patient with labial appliances.

Recommendations for further research

Parallel randomized controlled trials are needed in order to robustly compare lingual and labial orthodontic fixed appliances and should be preferred over non-randomized design, as clear evidence of bias was seen from the latter. These should ideally follow the CONSORT statement (58), be performed from multiple independent research centers and focus on long term outcomes pertaining to the completion of orthodontic treatment, possibly including the retention period. Primary focus should be thrown into objective measurements of therapeutic effects (like patient satisfaction and quality of life, the quality of final occlusion measured with the American Board of Orthodontics Objective Grading System, treatment duration, and relapse) or adverse effects (including root resorption, white spot lesions, gingival recessions, oral pain, oral discomfort, functional impairment, and cost of treatment).

Acknowledgements – We would like to thank Monique H. van der Veen (Preventive Dentistry, Academic Centre for Dentistry Amsterdam, Amsterdam, The Netherlands) for providing the raw data of their trial, and Dr. Yıldız Öztürk (Department of Orthodontics, Istanbul University, Istanbul, Turkey) and Dr. Tarek Z. Khattab (Department of Orthodontics, Hama University, Hama, Syria) for providing clarifications on the nature of their trial.

Conflicts of interest – No specific funding for this study. All authors confirm that there is no actual or perceived financial or other conflict of interest regarding this study. Mr. Papageorgiou is receiving funds by the Clinical Research Unit 208 (University of Bonn, Bonn, Germany).

References

1. KEIM RG, GOTTLIEB EL, NELSON AH, VOGELS DS 3RD. 2013 JCO Orthodontic Practice Study. Part 1: trends. *J Clin Orthod* 2013; **47**: 661–680.
2. ROSVALL MD, FIELDS HW, ZIUCHKOVSKI J, ROSENSTIEL SF, JOHNSTON WM. Attractiveness, acceptability, and value of orthodontic appliances. *Am J Orthod Dentofac Orthop* 2009; **135**: 276.e1–12.
3. FUJITA K. New orthodontic treatment with lingual bracket and mushroom arch wire appliance. *Am J Orthod* 1979; **76**: 657–675.
4. GERON S, SHPACK N, KANDOS S, DAVIDOVITCH M, VARDIMON AD. Anchorage loss—a multifactorial response. *Angle Orthod* 2003; **73**: 730–737.
5. YE L, KULA KS. Status of lingual orthodontics. *World J Orthod* 2006; 7: 361-8.
6. PAPAGEORGIOU SN, KONSTANTINIDIS I, PAPADOPOULOU K, JÄGER A, BOURAUEL C. Clinical effects of pre-adjusted edgewise orthodontic brackets: a systematic review and meta-analysis. *Eur J Orthod* 2014; **36**: 350–363.
7. PAPAGEORGIOU SN, KONSTANTINIDIS I, PAPADOPOULOU K, JÄGER A, BOURAUEL C. A systematic review and meta-analysis of experimental clinical evidence on initial aligning archwires and archwire sequences. *Orthod Craniofac Res* 2014; **17**: 197–215.
8. LONG H, ZHOU Y, PYAKUREL U, LIAO L, JIAN F, XUE J, YE N, YANG X, WANG Y, LAI W. Comparison of adverse effects between lingual and labial orthodontic treatment. *Angle Orthod* 2013*;* **83**: 1066–1073.
9. MISTAKIDIS I, KATIB H, VASILAKOS G, KLOUKOS D, GKANTIDIS N. Clinical outcomes of lingual orthodontic treatment: a systematic review. *Eur J Orthod* 2015 [Epub ahead of print]. DOI: http://dx.doi.org/10.1093/ejo/cjv061.
10. PAPAGEORGIOU SN, XAVIER GM, COBOURNE MT. Basic study design influences the results of orthodontic clinical investigations. *J Clin Epidemiol* 2015 [Epub ahead of print]. DOI: 10.1016/j.jclinepi.2015.03.008.
11. PAPAGEORGIOU SN. Meta-analysis for orthodontists: Part I--How to choose effect measure and statistical model. *J Orthod* 2014; **41**: 317–326.
12. PAPAGEORGIOU SN. Meta-analysis for orthodontists: Part II--Is all that glitters gold? *J Orthod* 2014; **41**: 327–336.

13. GUYATT GH, OXMAN AD, SCHÜNEMANN HJ, TUGWELL P, KNOTTNERUS A. GRADE guidelines: a new series of articles in the J Clin Epidemiol. *J Clin Epidemiol* 2011; **64**: 380–382.

14. HIGGINS JP, THOMPSON SG, SPIEGELHALTER DJ. A re-evaluation of random-effects meta-analysis. *J R Stat Soc Ser A Stat Soc* 2009; **172**: 137–159.

15. SHAMSEER L, MOHER D, CLARKE M, GHERSI D, LIBERATI A, PETTICREW M, SHEKELLE P, STEWART LA; PRISMA-P GROUP. Preferred Reporting Items for Systematic Review and Meta-Analysis Protocols (PRISMA-P) 2015: elaboration and explanation. *BMJ* 2015; **349**: g7647.

16. HIGGINS JPT, GREEN S, eds. Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 [updated March 2011]. The Cochrane Collaboration, 2011. Available at: www.cochrane-handbook.org. Accessed October 15th 2015.

17. LIBERATI A, ALTMAN DG, TETZLAFF J, MULROW C, GØTZSCHE PC, IOANNIDIS JP, CLARKE M, DEVEREAUX PJ, KLEIJNEN J, MOHER D. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. *J Clin Epidemiol* 2009; **62**: e1–e34.

18. PAPAGEORGIOU SN, KEILIG L, HASAN I, JÄGER A, BOURAUEL C. Effect of material variation on the biomechanical behaviour of orthodontic fixed appliances: a finite element analysis. *Eur J Orthod* 2015 [Epub ahead of print]. DOI: http://dx.doi.org/10.1093/ejo/cjv050.

19. PAPAGEORGIOU SN, TSIRANIDOU E, ANTONOGLOU GN, DESCHNER J, JÄGER A. Choice of effect measure for meta-analyses of dichotomous outcomes influenced the identified heterogeneity and direction of small-study effects. *J Clin Epidemiol* 2015; **68**: 534–541.

20. CHINN S. A simple method for converting an odds ratio to effect size for use in meta-analysis. *Stat Med* 2000; **19**: 3127–3131.

21. HIGGINS JP, THOMPSON SG. Quantifying heterogeneity in a meta-analysis. *Stat Med* 2002; **21**: 1539–1558.

22. HIGGINS JP, THOMPSON SG, DEEKS JJ, ALTMAN DG. Measuring inconsistency in meta-analyses. *BMJ* 2003; **327**: 557–560.

23. IOANNIDIS JP, PATSOPOULOS NA, EVANGELOU E. Uncertainty in heterogeneity estimates in meta-analyses. *BMJ* 2007; **335**: 914–916.

24. ORSINI N, BOTTAI M, HIGGINS J, BUCHAN I. Heterogi: Stata module to quantify heterogeneity in a meta-analysis. Statistical Software Components 2006. www.EconPapers.repec.org/RePEc:boc:bocode:s449201 (15 August 2015, date last accessed).

25. IOANNIDIS JP. Interpretation of tests of heterogeneity and bias in meta-analysis. *J Eval Clin Pract* 2008; **14**: 951–957.

26. SLOAN J, SYMONDS T, VARGAS-CHANES D, FRIDLEY B. Practical guidelines for assessing the clinical significance of health-related quality of life changes within clinical trials. *Drug Inf J* 2006; **37**: 23–31.

27. KNAPP G, HARTUNG J. Improved tests for a random-effects meta-regression with a single covariate. *Stat Med* 2003; **22**: 2693–2710.

28. EGGER M, DAVEY SMITH G, SCHNEIDER M, MINDER C. Bias in meta-analysis detected by a simple, graphical test. *BMJ* 1997; **315**: 629–34.

29. CANIKLIOGLU C, OZTURK Y. Patient discomfort: a comparison between lingual and labial fixed appliances. *Angle Orthod* 2005; **75**: 86–91.

30. KHATTAB TZ, FARAH H, AL-SABBAGH R, HAJEER MY, HAJ-HAMED Y. Speech performance and oral impairments with lingual and labial orthodontic appliances in the first stage of fixed treatment. *Angle Orthod* 2013; **83**: 519–526.

31. KHATTAB TZ, HAJEER MY, FARAH H, AL-SABBAGH R. Maxillary dental arch changes following the leveling and alignment stage with lingual and labial orthodontic appliances: a preliminary report of a randomized controlled trial. *J Contemp Dent Pract* 2014; **15**: 561–566.

32. LOMBARDO L, ORTAN YO, GORGUN O, PANZA C, SCUZZO G, SICILIANI G. Changes in the oral environment after placement of lingual and labial orthodontic appliances. *Prog Orthod* 2013; **14**: 28.

33. RAI AK, GANESHKAR SV, ROZARIO JE. Parametric and nonparametric assessment of speech changes in labial and lingual orthodontics: A prospective study. *APOS Trends Orthod* 2013; **3**: 99–109.

34. RAI AK, ROZARIO JE, GANESHKAR SV. Comparison of speech performance in labial and lingual orthodontic patients: A prospective study. *Dent Res J (Isfahan)* 2014; **11**: 663–675.

35. SHALISH M, COOPER-KAZAZ R, IVGI I, CANETTI L, TSUR B, BACHAR E, CHAUSHU S. Adult patients' adjustability to orthodontic appliances. Part I: a comparison between Labial, Lingual, and Invisalign. *Eur J Orthod* 2012; **34**: 724–730.

36. SOLDANOVA M. [Comparison of treatment effectiveness lingual apparatus 2D and apparatus straight wire]. Olomouc: Palacký University, 2011; (MSc Thesis).

37. SOLDANOVA M, LESETICKY O, KOMARKOVA L, DOSTALOVA T, SMUTNY V, SPIDLEN M. Effectiveness of treatment of adult patients with the straightwire technique and the lingual two-dimensional appliance. *Eur J Orthod* 2012; **34**: 674–680.

38. VAN DER VEEN MH, ATTIN R, SCHWESTKA-POLLY R, WIECHMANN D. Caries outcomes after orthodontic treatment with fixed appliances: do lingual brackets make a difference? *Eur J Oral Sci* 2010; **118**: 298–303.

39. VENKATESH S, ROZARIO J, GANESHKAR SV, AJMERA S. Comparative evaluation of sagittal anchorage loss in lingual and labial appliances during space closure: A pilot study. *APOS Trends Orthod* 2015; **5**: 33–37.

40. WU AK, MCGRATH C, WONG RW, WIECHMANN D, RABIE AB. A comparison of pain experienced by patients treated with labial and lingual orthodontic appliances. *Eur J Orthod,* 2010; **32**: 403–407.

41. WU A, MCGRATH C, WONG RW, WIECHMANN D, RABIE AB. Comparison of oral impacts experienced by patients treated with labial or customized lingual fixed orthodontic appliances. *Am J Orthod Dentofac Orthop* 2011; **139**: 784–790.

42. SINCLAIR PM, CANNITO MF, GOATES LJ, SOLOMOS LF, ALEXANDER CM. Patient responses to lingual appliance. *J Clin Orthod* 1986; **20**: 396–404.

43. ARTUN JA. A post-treatment evaluation of multibonded lingual appliances in lingual orthodontics. *Eur J Orthod* 1987; **9**: 204–210.

44. ROTTER JB. Generalized expectancies for internal versus external control of reinforcement. *Psychol Monogr* 1966; **80**: 1–28.

45. JONES M, CHAN C. The pain and discomfort experienced during orthodontic treatment: a randomized controlled clinical trial of two initial aligning archwires. *Am J Orthod Dentofac Orthop* 1992; **102**: 373–381.

46. SERGEL HG, KLAGES U, ZENTNER A. Pain and discomfort during orthodontics treatment—causative factors and effects on compliance. *Am J Orthod Dentofac Orthop* 1998; **114**: 684–691.

47. HOHOFF A, SEIFERT S, FILLION D, STAMM T, HEINECKE A, EHMER U. Speech performance in lingual orthodontic patients measured by sonagraphy and auditive analysis. *Am J Orthod Dentofac Orthop* 2003; **123**: 146–152.

48. HOHOFF A, FILLION D, STAMM T, GODER G, SAUERLAND C, EHMER U. Oral comfort, function and hygiene in patients with lingual brackets. *Journal of Orofacial Orthopedics* 2003; **64**: 359–371.

49. KURZ C, BENNETT R. Extraction cases and the lingual appliance. *J Am Lingual Orthod Assoc* 1988; **3**: 10–13.

50. STAMM T, HOHOFF A, EHMER U. A subjective comparison of two lingual bracket systems. *Eur J Orthod* 2005; **27**: 420–426.

51. CHANG HS, WALSH LJ, FREER TJ. Enamel demineralization during orthodontic treatment. Aetiology and prevention. *Aust Dent J* 1997; **42**: 322–327.

52. BRITSE A, LAGERLÖF F. The diluting effect of saliva on the sucrose concentration in different parts of the human mouth after a mouth-rinse with sucrose. *Arch Oral Biol* 1987; **32**: 755–756.

53. JENSEN B, BRATTHALL D. A new method for the estimation of Streptococci mutans in human saliva. *J Dent Res* 1989; **68**: 468–471.

54. NISHIMURA M, ODA T, KARIYA N, MATSUMURA S, SHIMONO T. Using a caries activity test to predict caries risk in early childhood. *J Am Dent Assoc* 2008; **139**: 63–71.

55. ROSENBLOOM RG, TINANOFF N. Salivary Streptococcus mutans levels in patients before, during, and after orthodontic treatment. *Am J Orthod Dentofac Orthop* 1991; **100**: 35–37.

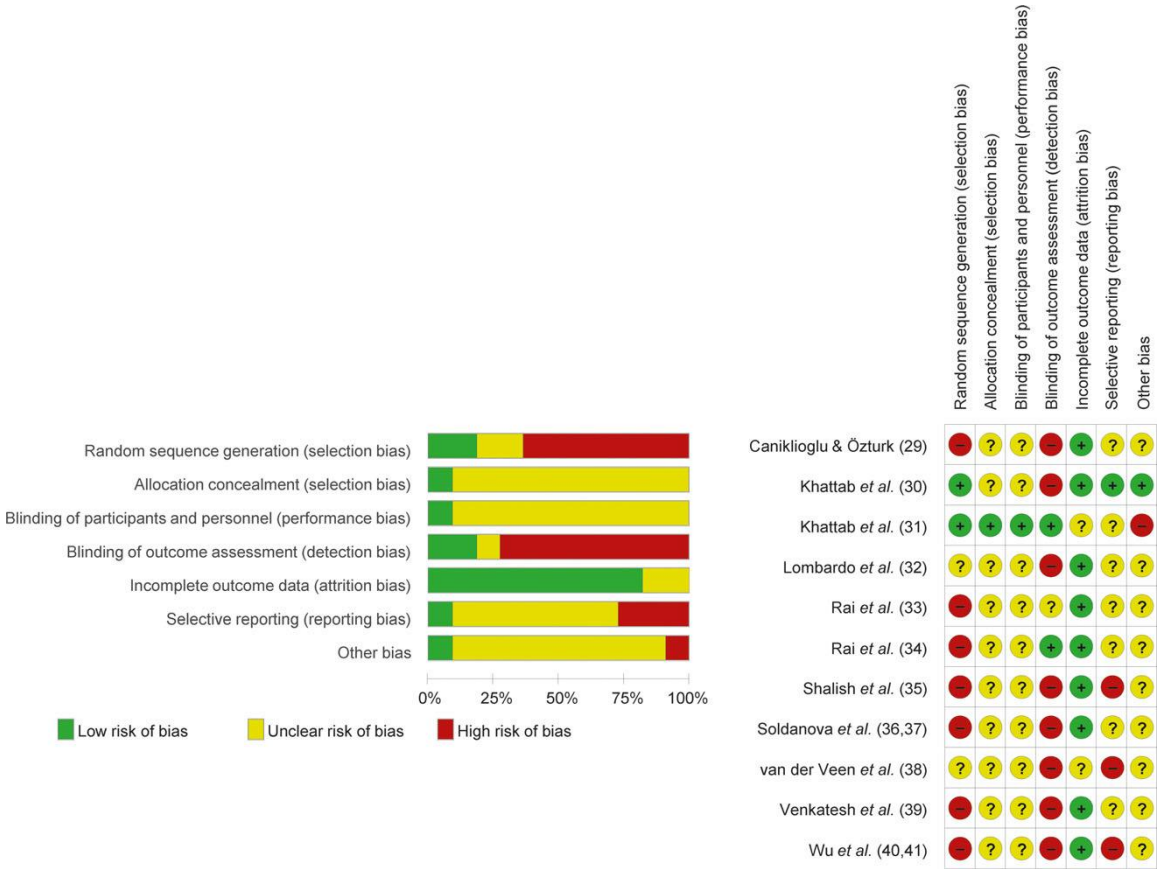
56. PAPAGEORGIOU SN, ANTONOGLOU GN, TSIRANIDOU E, JEPSEN S, JÄGER A. Bias and small-study effects influence treatment effect estimates: a meta-epidemiological study in oral medicine. *J Clin Epidemiol* 2014; **67**: 984–992.

57. PAPAGEORGIOU SN, KLOUKOS D, PETRIDIS H, PANDIS N. Publication of statistically significant research findings in prosthodontics & implant dentistry in the context of other dental specialties. *J Dent* 2015 [Epub ahead of print]. DOI: 10.1016/j.jdent.2015.08.005.

58. PANDIS N, FLEMING PS, HOPEWELL S, ALTMAN DG. The CONSORT Statement: Application within and adaptations for orthodontic trials. *Am J Orthod Dentofac Orthop* 2015; **147**: 663–679.

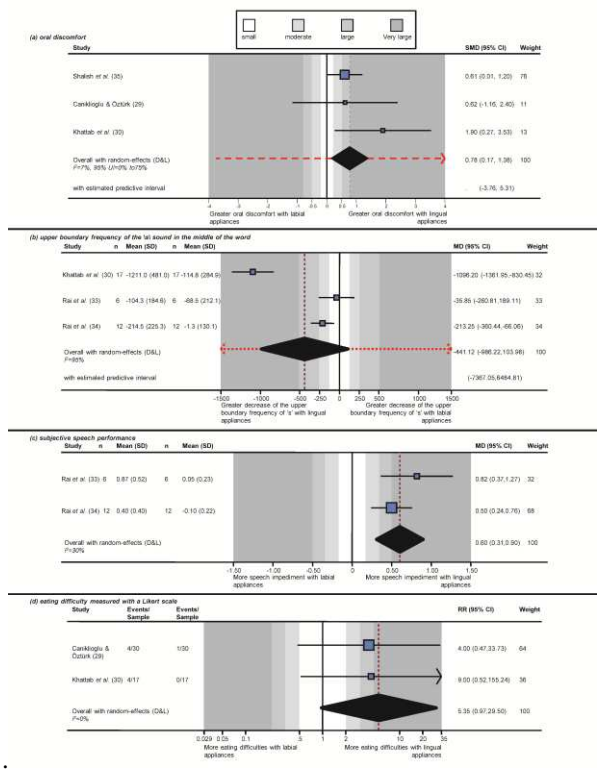
Figure legends

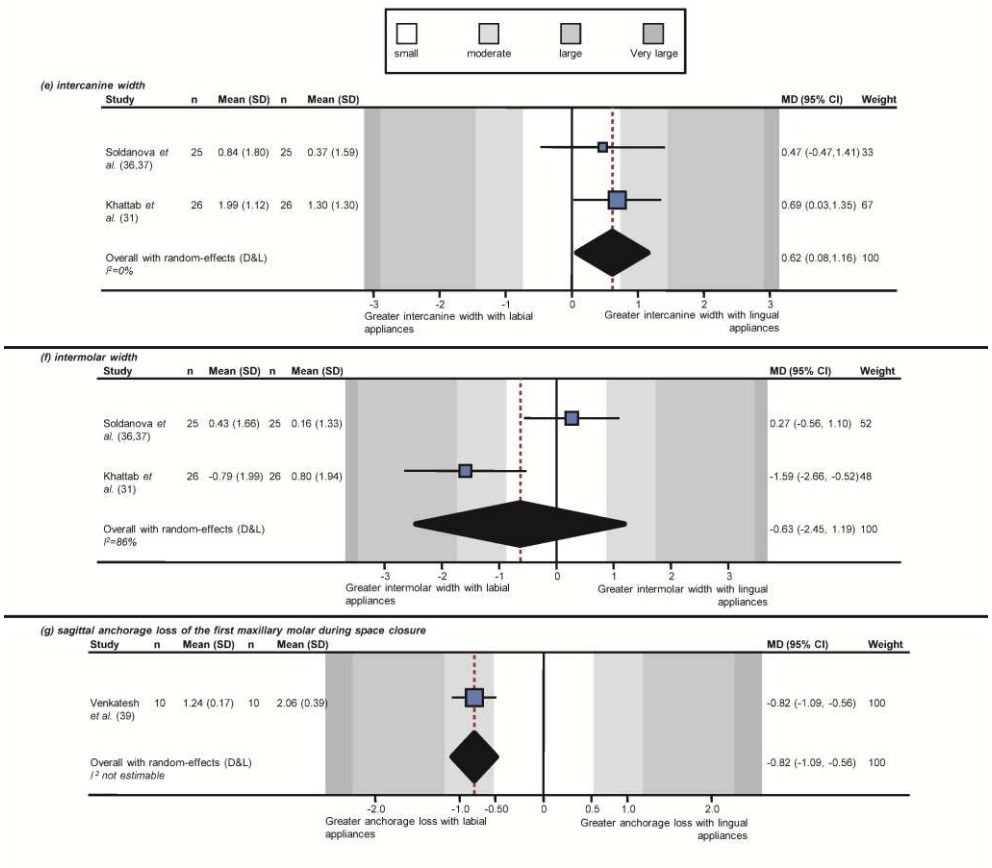
Fig. 1. Summary of the risk of bias of the trials included in this systematic review.



.

Fig. 2. Contour-enhanced forest plot of the treatment effects with lingual and labial appliances: (a), oral discomfort; (b), upper boundary frequency of the \s\ sound in the middle of the word; (c), subjective speech performance; (d), eating difficulty measured with a Likert scale; (e), intercanine width; (f), intermolar width; (g), sagittal anchorage loss of the first maxillary molar during space closure. Color contours indicate increasing effect magnitude from the middle to the ends of the forest plot: small effects (white), moderate effects (light grey), large effects (dark grey), and very large effects (darker grey).




small
moderate
large
Very large
(e) intercanine width
Study
n
Mean (SD)
n
Mean (SD)
MD (95% CI)
Weight
Soldanova et al. (36,37)
25
0.84 (1.80)
25
0.37 (1.59)
0.47 (-0.47,1.41) 33
Khattab et al. (31)
26
1.99 (1.12)
26
1.30 (1.30)
0.69 (0.03,1.35) 67
Overall with random-effects (D&L)
I²=0%
0.62 (0.08,1.16) 100
-3
-2
-1
0
1
2
3
Greater intercanine width with labial appliances
Greater intercanine width with lingual appliances
(f) intermolar width
Study
n
Mean (SD)
n
Mean (SD)
MD (95% CI)
Weight
Soldanova et al. (36,37)
25
0.43 (1.66)
25
0.16 (1.33)
0.27 (-0.56, 1.10) 52
Khattab et al. (31)
26
-0.79 (1.99)
26
0.80 (1.94)
-1.59 (-2.66, -0.52) 48
Overall with random-effects (D&L)
I²=86%
-0.63 (-2.45, 1.19) 100
-3
-2
-1
0
1
2
3
Greater intermolar width with labial appliances
Greater intermolar width with lingual appliances
(g) sagittal anchorage loss of the first maxillary molar during space closure
Study
n
Mean (SD)
n
Mean (SD)
MD (95% CI)
Weight
Venkatesh et al. (39)
10
1.24 (0.17)
10
2.06 (0.39)
-0.82 (-1.09, -0.56) 100
Overall with random-effects (D&L)
I² not estimable
-0.82 (-1.09, -0.56) 100
-2.0
-1.0
-0.50
0
0.5
1.0
2.0
Greater anchorage loss with labial appliances
Greater anchorage loss with lingual appliances

Tables

Table 1
Characteristics of the included trials

Trial	Design	Patients (M/F)	Mean age (yrs)	App	Product†	Slot size	Prescription	Prefab/ Indiv	Bonding	Outcomes	Conflict of interest
Caniklioglu 2005 (28)	pCCT University Turkey	60 (21/39)	17.9	Ling	Ormco 7th Generation (Ormco, Glendora, CA, USA)	NR		Prefab	Indirect (Max & Mand) with TARG+TR System®	Pat-rep problems: discomfort/ tongue-lip-cheek soreness/ eating, speech, and oral care difficulties/ adaptation period/ general problems (3 mos)	Not mentioned
				Lab	NR	NR	Roth	Prefab	Direct (Max & Mand)		
Khattab 2013[‡,¥] (29)	RCT University Syria	34 (13/21)	21.3	Ling	Stealth®(American Orthodontics, Sheboygan, Wisc)	0.022”	Roth	Prefab	Indirect (Max) with TARG+TR System®	Speech performance evaluated with auditive analysis / Pat-rep oral impairment using a Likert-scale: oral discomfort, speech impairment, mastication difficulties (Bef-Tx, 1 mo, 3 mos)	Not mentioned
				Lab	Mini Master Series (American Orthodontics, Sheboygan, Wisc)	0.022”	Roth	Prefab	Direct (Max)		
Khattab 2014[*,‡,¥] (30)	RCT University Syria	52 (20/32)	21.2	Ling	Stealth®(American Orthodontics, Sheboygan, Wisc)	0.022”	Roth	Prefab	Indirect (Max) with TARG+TR System®	Intercanine-/ interpremolar-/ intermolar width; arch length; amount of enamel reduction (Bef-Tx, after leveling/aligning)	None
				Lab	Mini Master Series (American Orthodontics, Sheboygan, Wisc)	0.022”	Roth	Prefab	Direct (Max)		
Lombardo 2013 (31)	RCT University/practice (?) Turkey/Italy (?)	20 (5/15)	20.8	Ling	STb brackets (Ormco Corporation, Glendora, CA, USA)	0.018”		Prefab	NR (Max & Mand)	Oral health parameters: DMFT, PI, GBI / salivary flow rate / salivary buffer capacity and pH / S. mutans count / Lactobacillus count (Bef-Tx, 0 wk, 4 wks, 8 wks)	None; funded by research grant
				Lab	(American Orthodontics, Sheboygan, Wisc; Roth prescription);	0.018”	Roth	Prefab	NR (Max & Mand)		
Rai 2013 (32)	pCCT; University; India	12 (NR)	NR	Ling	STb brackets (Ormco Corporation, Glendora, CA, USA)	-	-	Prefab	Indirect (Max & Mand) with TAD+BPD System®	Objective, semiobjective, and subjective speech performance (Bef-Tx, 1 d, 1 wk, 1 mo)	None
				Lab	MBT Versatile+ brackets (3M Unitek, Monrovia, Calif)	-	MBT	Prefab	Direct (Max & Mand)		
Rai 2014 (33)	pCCT; University; India	24 (11/13)	23.0	Ling	STb brackets (Ormco Corporation, Glendora, CA, USA)	-	-	Prefab	Indirect (Max & Mand) with TAD+BPD System®	Objective, semiobjective, and subjective speech performance (Bef-Tx, 1 d, 1 wk, 1 mo)	None

				Lab[#]	MBT Versatile+ brackets (3M Unitek, Monrovia, Calif)	-	MBT	Prefab	Direct (Max & Mand)		
Shalish 2012[$] (34)	pCCT University/practices Israel	47 (18/29)	NR	Ling	Incognito (3M-Unitek, Monrovia, Calif)	0.018”		Indiv	Indirect (Max & Mand)	Pat-rep health-related quality of life using a Likert-scale / Pat-rep pain intensity and analgesic consumption / number of days needed to achieve mild or no pain (days 1-7 & day 14 after appliance insertion)	Not mentioned
				Lab	(GAC International, Inc., Bohemia NY, USA or Ormco, Glendora, CA, USA)	0.022”		Prefab	Direct (Max & Mand)		
Soldanova 2011;2012 (35, 36)	pCCT University Czech Republic	50 (11/39)	31.0	Ling	2D brackets (Forestadent, St Louis, Missouri, USA);	NR		Prefab	(Max & Mand)	Treatment duration; Intercanine-, interpremolar-, intermolar width / arch length / cephalometric measurements for the sagittal and vertical position of the lower incisors (Bef-Tx, Aft-Tx)	No mention; funded by research grant
				Lab	Minitrim brackets (Dentaurum, Ispringen, Germany	0.018”	Roth	Prefab	(Max & Mand)		
van der Veen 2010 (37)	split-mouth (Max-Mand) RCT practice Germany	28 (NR)	15.3	Ling	Incognito (TOP Service for Lingualtechnik, Bad Essen, Germany)	NR		Indiv	Indirect (Max or Mand)	Number of white spot lesions / (Bef-Tx, After-Tx)	Financial interest in product
				Lab	Orthos (Ormco, Glendorra, CA, USA)	NR		Prefab	Direct (Max or Mand)		
Venkatesh 2015[@] (38)	pCCT; University; India	20 (NR)	20.0	Ling	STb brackets (Ormco Corporation, Glendora, CA, USA)	0.018”	MBT	Prefab	Indirect (Max & Mand) with TAD+BPD System®	Cephalometric sagittal anchorage loss of the first maxillary permanent molar (before and after space closure)	None
				Lab	Victory brackets (3M Unitek, Monrovia, Calif)	0.018”	MBT	Prefab	Direct (Max & Mand)		
Wu 2010; 2011 (39, 40)	pCCT University China	60 (20/40)	21.0	Ling	Incognito (TOP Service for Lingualtechnik, Bad Essen, Germany)	NR		Indiv	Indirect (Max & Mand)	Pat-rep pain experience / Pat-rep oral satisfaction: oral discomfort, mastication, speech disturbances, and social functioning using a VAS-scale / Pat-rep sleep disturbance, analgesic consumption, and timing of initial pain (1 wk, 1 mos, 3 mos)	Not mentioned; funded by research grant
				Lab	Mini-Diamond (Ormco, Orange, California, USA)	NR		Prefab			

M/F, male/female; yrs, years; App, appliance; pCCT, prospective non-randomized clinical trial; Max, maxilla; Man, mandible; NR, not reported; TARG-TR, torque angulation reference guide + thickness & rotation; Pat-rep, patient-reported; mos, months; RCT, randomized controlled trial; Bef-Tx, before treatment; M1, first permanent molar; SS, stainless steel; DMFT, Decayed, missing, and filled teeth; PI, plaque index; GBI, gingival bleeding index; wk, week; Aft-Tx, after treatment; VAS, visual analogue scale.

*lower arches in both LI and LA groups were treated with labial appliances.

[‡]0.5-1.0mm bite-ramps on lower M1

[#]Goshgarian transpalatal arch used on all patients with labial appliances

[@]Nickel-Titanium closed coil spring and powerchains were used on the upper and lower dentition, respectively, for space closure

[$]a third patient group treated with Invisalign is omitted.

[†]all appliances were conventionally-ligated; not self-ligated.

[¥]Including also data from communication with the trial’s authors.

Table 2

Results of the performed meta-analyses

Type	Source	Outcome (time)	n	Effect	95% CI	P	I^2
Discomfort	Questionnaire	Oral discomfort (0.5-3 months)	3	SMD=0.78	0.18,1.38*	0.012	7% (0%,75%)
Speech	Auditory analysis	Upper boundary frequency of the /s/ sound in the middle of the word (after insertion)	3	MD=-722.29	-1500.00,94.26**	0.083	97% (96%,68%)
	Auditory analysis	Upper boundary frequency of the /s/ sound in the middle of the word (1 week)	2	MD=-312.18	-600.98,-23.39	0.034	75% (-)
	Auditory analysis	Upper boundary frequency of the /s/ sound in the middle of the word (1 month)	3	MD=-441.12	-986.22,103.98***	0.113	95% (90%,97%)
	Auditory analysis	Upper boundary frequency of the /s/ sound in the start of the word (after insertion)	2	MD=-39.87	-195.65,115.91	0.616	0% (-)
	Auditory analysis	Upper boundary frequency of the /s/ sound in the start of the word (1 week)	2	MD=-216.93	-372.62,-61.25	0.006	0% (-)
	Auditory analysis	Upper boundary frequency of the /s/ sound in the start of the word (1 month)	2	MD=-120.22	-282.73,42.29	0.147	0% (-)
	Clinical examination	Speech performance by layperson (1 day)	2	MD=0.00	-0.22,0.21	0.989	-
	Clinical examination	Speech performance by layperson (1 week)	2	MD=0.82	0.58,1.05	<0.001	-
	Clinical examination	Speech performance by layperson (1 month)	2	MD=0.60	0.31,0.90	<0.001	30% (-)
	Questionnaire	Speech disturbance (3 months)	2	RR=8.9	1.15,68.69	0.036	0% (-)
Eating	Questionnaire	Eating problems (3 months)	2	RR=5.35	0.97,29.50	0.054	0% (-)
Dental effects	Model analysis	Intercanine width (after Tx)	2	MD=0.62	0.08,1.16	0.025	0% (-)
	Model analysis	Interpremolar width (after Tx)	2	MD=-1.47	-3.41,0.48	0.139	87% (-)
	Model analysis	Intermolar width (after Tx)	2	MD=-0.63	-2.45,1.19	0.499	86% (-)
	Model analysis	Arch length to molars (after Tx)	2	MD=-0.48	-1.75,0.79	0.461	70% (-)
	Cephalometric analysis	Sagittal anchorage loss of the upper first molar (after Tx)	1	MD=-0.82	-1.09,-0.56	<0.001	-

SMD, standarized mean differences; MD, mean differences; Tx, treatment; RR, relative risk.

* with 95% predictive interval of -3.76 to 5.31.

** with 95% predictive interval of -11163.32 Hz to 9718.14 Hz.

*** with 95% predictive interval of -7367.05 Hz to 6484.81 Hz.

Table 3

GRADE summary of findings table for the main outcomes of the systematic review

Patients: receiving orthodontic treatment
Settings: university clinics (Turkey, Syria, India, Czech Republic)
Intervention: lingual fixed appliances
Comparison: labial fixed appliances

Outcomes	Illustrative comparative effects (95% CI)		Patients (trials)	GRADE*	Effect
	Labial appliances	Lingual appliances			
Oral discomfort; patient-reported (0.5-3 months)	Assumed risk per 1000 patients †	Corresponding risk per 1000 patients †	141 (3)	Very low ----	SMD=0.78 (0.18,1.38); P<0.05
	529 patient per 1000	1648 more patients per 1000 (204 to 5937 more)			
Upper boundary frequency of the /s/ sound in the middle of the word; auditory analysis (1 month)	Assumed change	Corresponding change	70 (3)	Very low ----	MD=-441.12 (-986.22,103.98); P>0.05
	The upper boundary frequency decreased on average by 61.51 Hz in the labial groups (range -1.27 Hz to -114.80 Hz)	The mean upper boundary frequency in the lingual groups decreased by 441.12 Hz (95% CI: 986.22 Hz decrease to 103.98 Hz increase) compared to the labial groups.			
Speech performance; assessed by layperson on a 5-point Likert scale (5=worst) (1 month)	Assumed change	Corresponding change	36 (2)	Very low ----	MD=0.60 (0.31,0.90); P<0.05
	The Likert score decreased on average by 0.03 points in the labial groups (range -0.10 to 0.05 points)	The mean Likert score in the lingual groups increased by 0.60 point (95% CI: 0.31 to 0.90 point increase) compared to the labial groups.			
Eating difficulty; patient-reported (3 months)	Assumed risk per 1000 patients	Corresponding risk per 1000 patients	94 (2)	Very low ----	RR=5.35 (0.97,29.50); P>0.05
	17 patient per 1000 (0 to 30)	72 more patients per 1000 (1 fewer to 475 more)			
Intercanine width; from dental cast analysis (after treatment)	Assumed change	Corresponding change	102 (2)	Very low ----	MD=0.62 (0.08,1.16); P<0.05
	The intercanine width increased on average by 0.84 mm in the labial groups (range 0.37 mm to 1.30 mm)	The mean intercanine width in the lingual groups increased by 0.62 mm (95% CI: 0.08 mm to 1.16 mm increase) compared to the labial groups.			
Intermolar width; from dental cast analysis (after treatment)	Assumed change	Corresponding change	102 (2)	Very low ----	MD=-0.63 (-2.45,1.19); P>0.05
	The intermolar width increased on average by 0.48 mm in the labial groups (range 0.16 mm to 0.80 mm)	The mean intermolar width in the lingual groups decreased by 0.63 mm (95% CI: 2.45 mm decrease to 1.19 mm increase) compared to the labial groups.			
Sagittal anchorage loss of the upper first molar during space closure; cephalometric analysis (after treatment)	Assumed change	Corresponding change	20 (1)	Very low ----	MD=-0.82 (-1.09,-0.56); P<0.05
	The mean anchorage loss in the labial group was 2.06 mm.	The mean anchorage loss in the lingual group decreased by 0.82 mm (95% CI: 0.56 mm to 1.09 mm decrease) compared to the labial group.			

CI, confidence interval; SMD, standardized mean difference; MD, mean difference; RR, relative risk; Tx, treatment.
†Assumed and corresponding changes calculated from the randomized trial of Khattab et al. (29), which was judged to be most robust.
*All GRADE scores start from low, due to the inclusion of non-randomized trials and are further downgraded at least by one (see Appendix 6).

Supporting Information

Lingual versus labial fixed orthodontic appliances: systematic review and meta-analysis of treatment effects

PAPAGEORGIOU SN, GÖLZ L, JÄGER A, ELIADES T, BOURAUEL C
School of Dentistry, University of Bonn, Germany; and Faculty of Medicine, University of Zurich, Switzerland

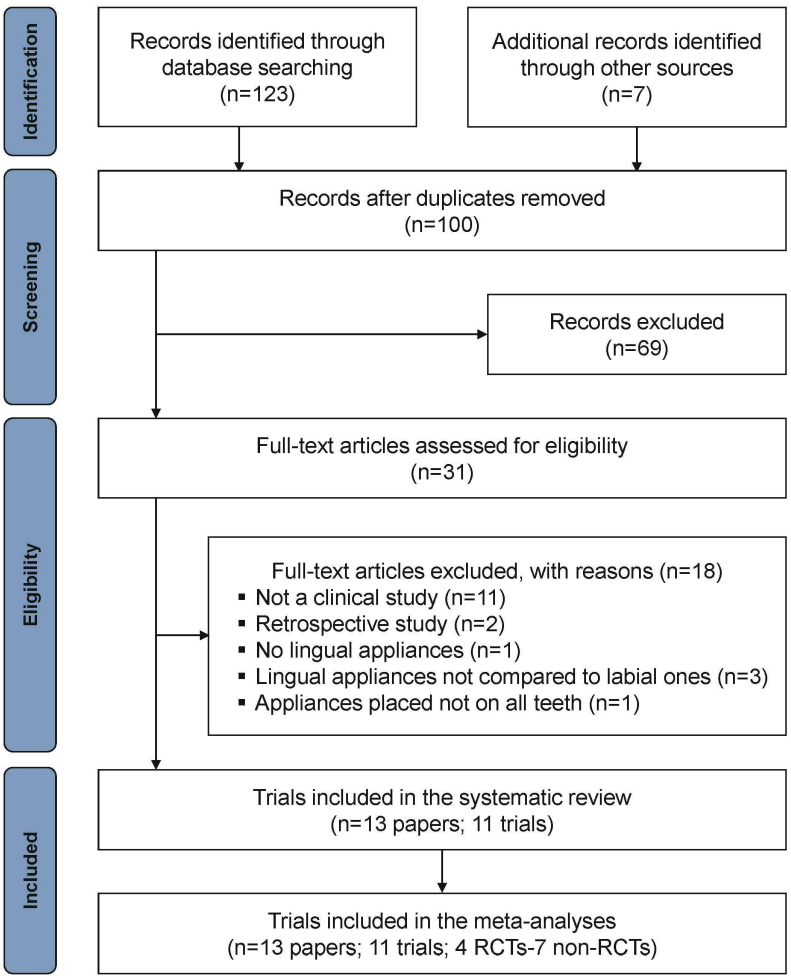


Fig. S1: Flow diagram for the identification and selection of studies

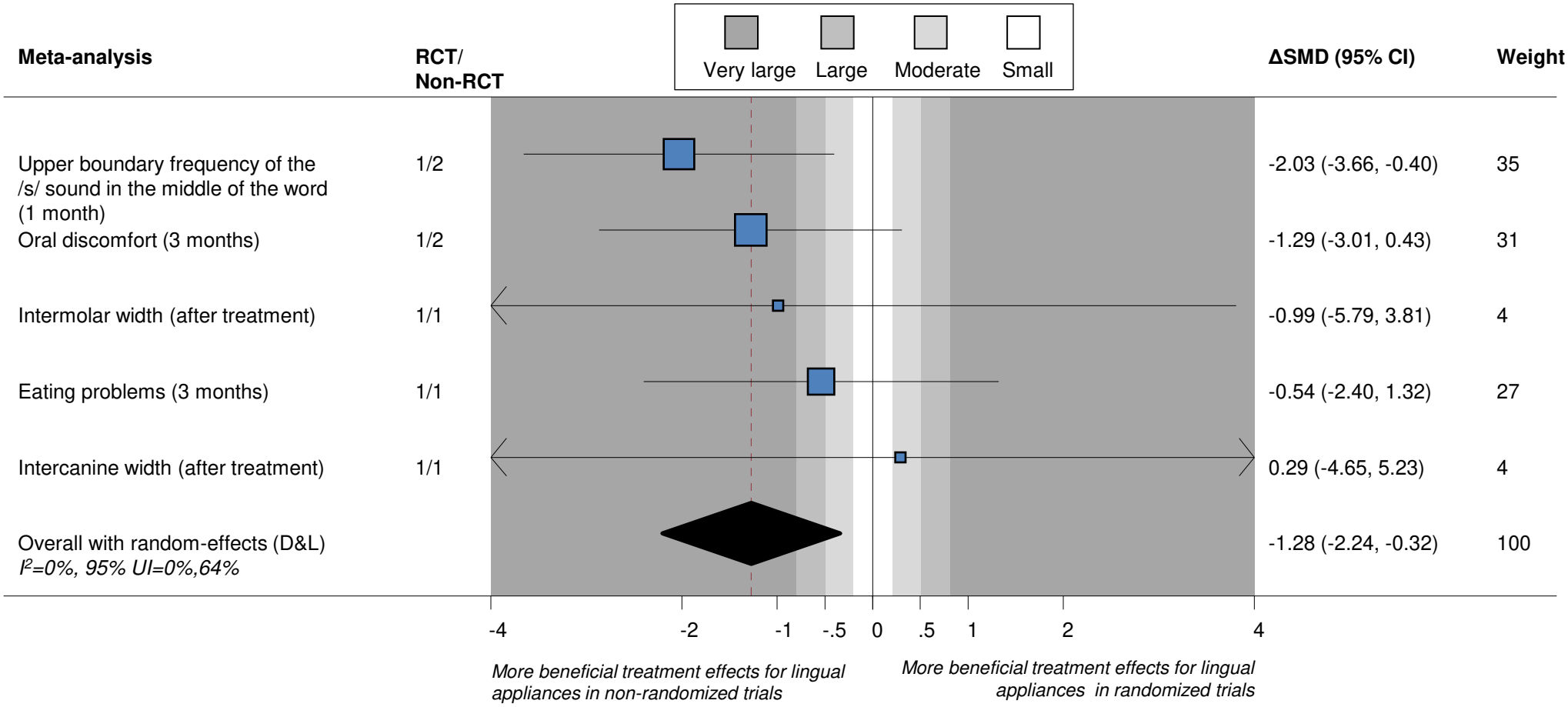


Fig. S2: Forest plot for the comparison of treatment effects from randomized and non-randomized trials. The difference in effects between randomized and non-randomized trials is expressed as difference in the standardized mean differences (ΔSMD) through random-effects meta-regression. Binary outcomes are appropriately converted to continuous and the ΔSMDs from all meta-analyses of Table S7 are pooled together with random-effects meta-analysis. RCT, randomized controlled trial; ΔSMD, as difference in the standardized mean differences; CI, confidence interval; D&L, DerSimonian & Laird method; UI, uncertainty interval.

Table S1

The electronic databases searched, the search strategy used, and the corresponding results (as of July 20th, 2015)

Database	Search Strategy	Limitations	Hits
MEDLINE searched through PubMed on July 20th, 2015 http://www.ncbi.nlm.nih.gov/pubmed/	orthodon* AND (lingual OR Incognito OR Harmony OR Magic OR iBraces OR "7th Generation" OR WIN OR STb OR 2D) AND labial AND (bracket* OR appliance*)	Clinical Trial; Comparative Study; Controlled Clinical Trial; Randomized Controlled Trial; Humans	57
Cochrane Database of Systematic Reviews searched on July 20th, 2015 http://onlinelibrary.wiley.com/cochranelibrary/search/	orthodon* AND (lingual OR Incognito OR Harmony OR Magic OR iBraces OR "7th Generation" OR WIN OR STb OR 2D) AND labial AND (bracket* OR appliance*)	-	1
Database of Abstracts of Reviews of Effects searched on July 20th, 2015 http://onlinelibrary.wiley.com/cochranelibrary/search/	orthodon* AND (lingual OR Incognito OR Harmony OR Magic OR iBraces OR "7th Generation" OR WIN OR STb OR 2D) AND labial AND (bracket* OR appliance*)	-	1
Cochrane Central Register of Controlled Trials searched on July 20th, 2015 http://onlinelibrary.wiley.com/cochranelibrary/search/	orthodon* AND (lingual OR Incognito OR Harmony OR Magic OR iBraces OR "7th Generation" OR WIN OR STb OR 2D) AND labial AND (bracket* OR appliance*)	-	10
Virtual Health Library searched on July 20th, 2015 http://regional.bvsalud.org/	orthodon* AND (lingual OR Incognito OR Harmony OR Magic OR iBraces OR "7th Generation" OR WIN OR STb OR 2D) AND labial AND (bracket* OR appliance*)	-	10
Web of Science searched on July 20th, 2015 https://isiknowledge.com/	orthodon* AND (lingual OR Incognito OR Harmony OR Magic OR iBraces OR "7th Generation" OR WIN OR STb OR 2D) AND labial AND (bracket* OR appliance*)	Dentistry, oral surgery, medicine	44
	Sum		123

Table S2

List of included and excluded studies, with the corresponding reasons

Paper	Decision
Almeida MR, Henriques JF, Almeida RR, Almeida-Pedrin RR, Ursi W. Treatment effects produced by the Bionator appliance. Comparison with an untreated Class II sample. Eur J Orthod 2004;26:65-72.	Excluded by title
Anehus-Pancherz M, Pancherz H. [The effect on chewing of treating distal bite with an activator]. J Orofac Orthop 1989;50:392-405.	Excluded by title
Atack N, Harradine N, Sandy JR, Ireland AJ. Which way forward? Fixed or removable lower retainers. Angle Orthod 2007;77:954-9.	Excluded by title
Atsu SS, Gelgor IE, Sahin V. Effects of silica coating and silane surface conditioning on the bond strength of metal and ceramic brackets to enamel. Angle Orthod 2006;76:857-62.	Excluded by title
Basciftci FA, Uysal T, Buyukerkmen A, Sari Z. The effects of activator treatment on the craniofacial structures of Class II division 1 patients. Eur J Orthod 2003;25:87-93.	Excluded by title
Carlstedt K, Henningsson G, Dahllof G. A four-year longitudinal study of palatal plate therapy in children with Down syndrome: effects on oral motor function, articulation and communication preferences. Acta Odontol Scand 2003;61:39-46.	Excluded by title
Carlstedt K, Henningsson G, McAllister A, Dahllof G. Long-term effects of palatal plate therapy on oral motor function in children with Down syndrome evaluated by video registration. Acta Odontol Scand 2001;59:63-8.	Excluded by title
Chumak L, Galil KA, Way DC, Johnson LN, Hunter WS. An in vitro investigation of lingual bonding. Am J Orthod Dentofac Orthop 1989;95:20-8.	Excluded by title
De Almeida MR, Henriques JF, Ursi W. Comparative study of the Frankel (FR-2) and bionator appliances in the treatment of Class II malocclusion. Am J Orthod Dentofac Orthop 2002;121:458-66.	Excluded by title
de Cuebas JO. Nonsurgical treatment of a skeletal vertical discrepancy with a significant open bite. Am J Orthod Dentofac Orthop 1997;112:124-31.	Excluded by title
Dittmer MP, Demling AP, Borchers L, Stiesch M, Kohorst P, Schwestka-Polly R. Tensile properties of orthodontic elastomeric chains. J Orofac Orthop 2010;71:330-8.	Excluded by title
Farquhar RB. Direct bonding comparing a polyacrylic acid and a phosphoric acid technique. Am J Orthod Dentofac Orthop 1986;90:187-94.	Excluded by title
Fritz U, Diedrich P, Wiechmann D. Lingual technique--patients' characteristics, motivation and acceptance. Interpretation of a retrospective survey. J Orofac Orthop 2002;63:227-33.	Excluded by title
Gianelly AA. Leeway space and the resolution of crowding in the mixed dentition. Semin Orthod 1995;1:188-94.	Excluded by title
Gisel EG, Schwartz S, Petryk A, Clarke D, Haberfellner H. "Whole body" mobility after one year of intraoral appliance therapy in children with cerebral palsy and moderate eating impairment. Dysphagia 2000;15:226-35.	Excluded by title
Harazaki M, Isshiki Y. Soft laser irradiation effects on pain reduction in orthodontic treatment. The Bulletin of Tokyo Dental College 1997;38:291-5.	Excluded by title
Hohoff A, Ehmer U. Short-term and long-term results after early treatment with the Castillo Morales stimulating plate. A longitudinal study. J Orofac Orthop 1999;60:2-12.	Excluded by title
Honme Y, Motoyoshi M, Shinohara A, Shigeeda T, Shimizu N. Efficient palatal expansion with a quadhelix appliance: an in vitro study using an experimental dental arch model. Eur J Orthod 2012;34:442-6.	Excluded by title
Jost-Brinkmann PG, Tanne K, Sakuda M, Miethke RR. [A FEM study for the biomechanical comparison of labial and palatal force application on the upper incisors. Finite element method]. J Orofac Orthop 1993;54:76-82.	Excluded by title
Kannabiran P, Thirukonda GJ, Mahendra L. The crown angulations and inclinations in Dravidian population with normal occlusion. Indian J Dent Res 2012;23:53-8.	Excluded by title
Kinzinger G, Fritz U, Diedrich P. Various anchorage approaches in unilateral mandibular molar distalization using a fixed lingual arch appliance. J Orofac Orthop 2004;65:137-49.	Excluded by title
Kravitz ND, Kusnoto B, BeGole E, Obrez A, Agran B. How well does Invisalign work? A prospective clinical study evaluating the efficacy of tooth movement with Invisalign. Am J Orthod Dentofac Orthop 2009;135:27-35.	Excluded by title

Liang W, Rong Q, Lin J, Xu B. Torque control of the maxillary incisors in lingual and labial orthodontics: a 3-dimensional finite element analysis. Am J Orthod Dentofac Orthop 2009;135:316-22.	Excluded by title
Lombardo L, Scuzzo G, Arreghini A, Gorgun O, Ortan YO, Siciliani G. 3D FEM comparison of lingual and labial orthodontics in en masse retraction. Prog Orthod 2014;15:38.	Excluded by title
Lombardo L, Stefanoni F, Mollica F, Laura A, Scuzzo G, Siciliani G. Three-dimensional finite-element analysis of a central lower incisor under labial and lingual loads. Prog Orthod 2012;13:154-63.	Excluded by title
Lombardo L, Wierusz W, Toscano D, Lapenta R, Kaplan A, Siciliani G. Frictional resistance exerted by different lingual and labial brackets: an in vitro study. Prog Orthod 2013;14:37.	Excluded by title
Matsuda A, Suda N, Motohashi N, Tsuji M, Ohyama K. Skeletal characteristics and treatment outcome of five patients with Robin sequence. Angle Orthod 2006;76:898-908.	Excluded by title
Moran KI. Relative wire stiffness due to lingual versus labial interbracket distance. Am J Orthod Dentofac Orthop 1987;92:24-32.	Excluded by title
Ngan P, Hagg U, Yiu C, Merwin D, Wei SH. Treatment response to maxillary expansion and protraction. Eur J Orthod 1996;18:151-68.	Excluded by title
Okay C, Gulsen A, Keykubat A, Ucem TT, Yuksel S. A comparison of the effects of 2 mandibular anchorage systems used with a 3-dimensional bimetric maxillary distalizing arch. World J Orthod 2006;7:125-33.	Excluded by title
Opitz C, Muhler G, Bloch I, Schenk HJ. [A contribution to the controversial discussion on a preoperative orthodontic treatment for infants with unilateral cheilognathopalatoschisis]. J Orofac Orthop 1992;53:330-7.	Excluded by title
Ozturk Ortan Y, Yurdakuloglu Arslan T, Aydemir B. A comparative in vitro study of frictional resistance between lingual brackets and stainless steel archwires. Eur J Orthod 2012;34:119-25.	Excluded by title
Park JH, Lee YK, Lim BS, Kim CW. Frictional forces between lingual brackets and archwires measured by a friction tester. Angle Orthod 2004;74:816-24.	Excluded by title
Phillips C, Medland WH, Fields HW, Jr., Proffit WR, White RP, Jr. Stability of surgical maxillary expansion. Int J Adult Orthodon Orthognath Surg 1992;7:139-46.	Excluded by title
Quintao C, Helena I, Brunharo VP, Menezes RC, Almeida MA. Soft tissue facial profile changes following functional appliance therapy. Eur J Orthod 2006;28:35-41.	Excluded by title
Schuster G, Giese R. Retrospective clinical investigation of the impact of early treatment of children with Down's syndrome according to Castillo-Morales. J Orofac Orthop 2001;62:255-63.	Excluded by title
Showkatbakhsh R, Jamilian A, Taban T, Golrokh M. The effects of face mask and tongue appliance on maxillary deficiency in growing patients: a randomized clinical trial. Prog Orthod 2012;13:266-72.	Excluded by title
Sifakakis I, Pandis N, Makou M, Katsaros C, Eliades T, Bourauel C. A comparative assessment of forces and moments generated by lingual and conventional brackets. Eur J Orthod 2013;35:82-6.	Excluded by title
Smithpeter J, Covell D, Jr. Relapse of anterior open bites treated with orthodontic appliances with and without orofacial myofunctional therapy. Am J Orthod Dentofac Orthop 2010;137:605-14.	Excluded by title
So LL. Effects of reverse headgear treatment on sagittal correction in girls born with unilateral complete cleft lip and cleft palate--skeletal and dental changes. Am J Orthod Dentofac Orthop 1996;109:140-7.	Excluded by title
Sung SJ, Baik HS, Moon YS, Yu HS, Cho YS. A comparative evaluation of different compensating curves in the lingual and labial techniques using 3D FEM. Am J Orthod Dentofac Orthop 2003;123:441-50.	Excluded by title
Wang B, Shen G, Fang B, Yu H, Wu Y, Sun L. Augmented corticotomy-assisted surgical orthodontics decompensates lower incisors in Class III malocclusion patients. J Oral Maxillofac Surg 2014;72:596-602.	Excluded by title
Ahrens DG, Shapira Y, Kuftinec MM. An approach to rotational relapse. Am J Orthod 1981;80:83-91.	Excluded by title
Almeida MR, Henriques JF, Ursi W. Comparative study of the Fränkel (FR-2) and bionator appliances in the treatment of Class II malocclusion. Am J Orthod Dentofac Orthop 2002;5:458-66.	Excluded by title
Yu Y, Sun J, Lai W, Wu T, Koshy S, Shi Z. Interventions for managing relapse of the lower front teeth after orthodontic treatment. Cochrane Database Syst Rev 2013;9:CD008734.	Excluded by title
Atsü S, Çatalba, amp, Bülent, Gelgör IE. Effects of silica coating and silane surface conditioning on the bond strength of rebonded metal and ceramic brackets. J Appl Oral Sci 2011;19:233-9.	Excluded by title
Echarri P, Pedernera M. [Updated in the bracket positioning in the indirect bonding techniqueu]. Ortodoncia 2013;76:20-7.	Excluded by title

Menezes CCd. [Labial movement of the mandibular incisors and tomographic characteristics of the labial and lingual alveolar bone]. 2014; Doctoral thesis, Faculdade de Odontologia de Bauru.	Excluded by title
Piacenza A, Mondino N. [Another alternative in orthodontics. Spring retainers (elastic retainer)]. Claves Odontol 1995;2:8-9.	Excluded by title
Reis AC, Capelozza Filho L, Ozawa TO, Cavassan AdO. [Evaluation of tipping and inclination of teeth in young adults with complete bilateral cleft lip and palate]. Rev Dent Press Ortodon Ortopedi Facial 2008;13:113-23.	Excluded by title
Rothier EKC, Vilella OdV. [Technique to make temporary esthetic pontics for orthodontic appliances]. Rev Bras Odontol 2012;69:258-9.	Excluded by title
Chung K-R, Kim S-H, Lee B-S. Speedy surgical-orthodontic treatment with temporary anchorage devices as an alternative to orthognathic surgery. Am J Orthod Dentofacial Orthop 2009;135:787-98.	Excluded by title
Chung K-R, Kook Y-A, Kim S-H, Mo S-S, Jung J-A. Class II malocclusion treated by combining a lingual retractor and a palatal plate. Am J Orthod Dentofacial Orthop 2008;133:112-23.	Excluded by title
Geron S, Shpack N, Kandos S, Davidovitch M, Vardimon AD. Anchorage loss - A multifactorial response. Angle Orthod 2003;73:730-7.	Excluded by title; known to be retrospective
Janson G, de Souza JEP, Alves FD, Andrade P, Nakamura A, de Freitas MR, et al. Extreme dentoalveolar compensation in the treatment of Class III malocclusion. Am J Orthod Dentofacial Orthop 2005;128:787-94.	Excluded by title
Kawakami M, Miyawaki S, Noguchi H, Kirita T. Screw-type implants used as anchorage for lingual orthodontic mechanics: A case of bimaxillary protrusion with second premolar extraction. Angle Orthod 2004;74:715-9.	Excluded by title
Levin L, Samorodnitzky-Naveh GR, Machtei EE. The Association of Orthodontic Treatment and Fixed Retainers With Gingival Health. J Periodontol 2008;79:2087-92.	Excluded by title
Matsuda A, Suda N, Motohashi N, Tsuji M, Ohyama K. Skeletal characteristics and treatment outcome of five patients with Robin sequence cases. Angle Orthod 2006;76:898-908.	Excluded by title
Miyawaki S, Yasuhara M, Koh Y. Discomfort caused by bonded lingual orthodontic appliances in adult patients as examined by retrospective questionnaire. Am J Orthod Dentofacial Orthop 1999;115:83-8.	Excluded by title
Mo S-S, Kim S-H, Sung S-J, Chung K-R, Chun Y-S, Kook Y-A, et al. Factors controlling anterior torque during C-implant-dependent en-masse retraction without posterior appliances. Am J Orthod Dentofacial Orthop 2011;140:72-80.	Excluded by title
Nihara J, Gielo-Perczak K, Cardinal L, Saito I, Nanda R, Uribe F. Finite element analysis of mandibular molar protraction mechanics using miniscrews. Eur J Orthod 2015;37:95-100.	Excluded by title
Oesterle LJ, Shellhart WC. Bracket bond strength with transillumination of a light-activated orthodontic adhesive. Angle Orthod 2001;71:307-11.	Excluded by title
Phan X, Ling PH. Clinical limitations of invisalign. J Can Dent Assoc 2007;73:263-6.	Excluded by title
Wahl N. Orthodontics in 3 millennia. Chapter 16: Late 20th-century fixed appliances. Am J Orthod Dentofacial Orthop 2008;134:827-30.	Excluded by title
Wehrbein H, Bauer W, Diedrich P. Mandibular incisors alveolar bone, and symphysis after orthodontic treatment. A retrospective study. Am J Orthod Dentofacial Orthop 1996;110:239-46.	Excluded by title
Wiechmann D, Schwestka-Polly R, Hohoff A. Herbst appliance in lingual orthodontics. Am J Orthod Dentofacial Orthop 2008;134:439-46.	Excluded by title
Yanagita T, Nakamura M, Kawanabe N, Yamashiro T. Class II malocclusion with complex problems treated with a novel combination of lingual orthodontic appliances and lingual arches. Am J Orthod Dentofacial Orthop 2014;146:98-107.	Excluded by title
Kim SH, Park SB, Yang HC. Three-dimensional finite element analysis of the bracket positioning plane in lingual orthodontics. Korean J Orthod 2006;36:30-44.	Excluded by title
Mah YJ, Sohn H-K, Choi B-J, Lee J-H, Kim SO. Orthodontic traction of horizontally erupted lower lateral incisor on the lingual side. J Korean Acad Pediatr Dent 2010;37:117-23.	Excluded by title
Demling A, Dittmer MP, Schwestka-Polly R. Comparative analysis of slot dimension in lingual bracket systems. Head Face Med 2009;5:27.	Excluded by abstract; not a clinical study
Fuck LM, Wiechmann D, Drescher D. Comparison of the initial orthodontic force systems produced by a new lingual bracket system and a straight-wire appliance. J Orofac Orthop 2005;66:363-76.	Excluded by abstract; not a clinical study
Hugo A, Reyneke JP, Weber ZJ. Lingual orthodontics and orthognathic surgery. Int J Adult Orthodon Orthognath Surg 2000;15:153-62.	Excluded by abstract; not a clinical study

Shpack N, Geron S, Floris I, Davidovitch M, Brosh T, Vardimon AD. Bracket placement in lingual vs labial systems and direct vs indirect bonding. Angle Orthod 2007;77:509-17.	Excluded by abstract; not a clinical study
Ugur T, Yukay F. Normal faciolingual inclinations of tooth crowns compared with treatment groups of standard and pretorqued brackets. Am J Orthod Dentofac Orthop 1997;112:50-7.	Excluded by abstract; not a clinical study
Long H, Zhou Y, Pyakurel U, Liao L, Jian F, Xue J, Ye N, Yang X, Wang Y, Lai W. Comparison of adverse effects between lingual and labial orthodontic treatment. Angle Orthod 2013;83:1066-73.	Excluded by abstract; not a clinical study
Echarri P. [Correction of the anterior deep bite and anterior open bite with the double-arch technique and with labial or lingual brackets]. Ortod Esp 2003;43:240-51.	Excluded by abstract; not a clinical study
Galvão MCdS, Maltagliati LÁ, Sannomiya EK, Bommarito S. [Discomfort caused by lingual orthodontic appliance bonding versus labial]. Ortodontia 2008;41:19-24.	Excluded by abstract; not a clinical study
Monini AdC, Gandini Júnior LG, Gandini MREAS, Figueiredo JFBd. [Biomechanical differences between lingual and labial orthodontics]. Rev Dent Press Ortodon Ortopedi Facial 2008;13:92-100.	Excluded by abstract; not a clinical study
Geron S, Romano R, Brosh T. Vertical forces in labial and lingual orthodontics applied on maxillary incisors - A theoretical approach. Angle Orthod 2004;74:195-201.	Excluded by abstract; not a clinical study
Major TW, Carey JP, Nobes DS, Heo G, Melenka GW, Major PW. An investigation into the mechanical characteristics of select self-ligated brackets at a series of clinically relevant maximum torquing angles: loading and unloading curves and bracket deformation. Eur J Orthod 2013;35:719-29.	Excluded by abstract; not a clinical study
Looi LK, Mills JR. The effect of two contrasting forms of orthodontic treatment on the facial profile. Am J Orthod 1986;89:507-17.	Excluded by abstract; no lingual appliances
Lossdoerfer S, Schwestka-Polly R, Wiechmann D. Control of lower incisor inclination with a completely customized lingual appliance for dentoalveolar compensation of class III malocclusion. J Orofac Orthop 2013;74:381-96.	Excluded by abstract; lingual appliances not compared to labial appliances
Dalessandri D, Lazzaroni E, Migliorati M, Piancino MG, Tonni I, Bonetti S. Self-ligating fully customized lingual appliance and chair-time reduction: a typodont study followed by a randomized clinical trial. Eur J Orthod 2013;35:758-65.	Excluded by full-text; lingual appliances not compared to labial appliances
Demling A, Demling C, Schwestka-Polly R, Stiesch M, Heuer W. Short-term influence of lingual orthodontic therapy on microbial parameters and periodontal status. A preliminary study. Angle Orthod 2010;80:480-4.	Excluded by full-text; lingual appliances not compared to labial appliances
Sfondrini MF, Debiaggi M, Zara F, Brerra R, Comelli M, Bianchi M, Pollone SR, Scribante A. Influence of lingual bracket position on microbial and periodontal parameters in vivo. J Appl Oral Sci 2012;20:357-61.	Excluded by full-text; partial fixed-appliance (brackets only on selected teeth and no wire)
Deguchi T, Terao F, Aonuma T, Kataoka T, Sugawara Y, Yamashiro T, et al. Outcome assessment of lingual and labial appliances compared with cephalometric analysis, peer assessment rating, and objective grading system in Angle Class II extraction cases. Angle Orthod 2015;85:400-7.	Excluded by full-text; retrospective clinical trial
Gorman JC, Smith RJ. Comparison of treatment effects with labial and lingual fixed appliances. Am J Orthod Dentofac Orthop 1991;99:202-9.	Excluded by full-text; retrospective clinical trial
Caniklioglu C, Ozturk Y. Patient discomfort: a comparison between lingual and labial fixed appliances. Angle Orthod 2005;75:86-91.	Included
Khattab TZ, Farah H, Al-Sabbagh R, Hajeer MY, Haj-Hamed Y. Speech performance and oral impairments with lingual and labial orthodontic appliances in the first stage of fixed treatment. Angle Orthod 2013;83:519-26.	Included
Khattab TZ, Hajeer MY, Farah H, Al-Sabbagh R. Maxillary dental arch changes following the leveling and alignment stage with lingual and labial orthodontic appliances: a preliminary report of a randomized controlled trial. J Contemp Dent Pract 2014;15:561-6.	Included
Lombardo L, Ortan YO, Gorgun O, Panza C, Scuzzo G, Siciliani G. Changes in the oral environment after placement of lingual and labial orthodontic appliances. Prog Orthod 2013;14:28.	Included
Rai AK, Ganeshkar SV, Rozario JE. Parametric and nonparametric assessment of speech changes in labial and lingual orthodontics: A prospective study. APOS Trends Orthod 2013;3:99-109.	Included

Rai AK, Rozario JE, Ganeshkar SV. Comparison of speech performance in labial and lingual orthodontic patients: A prospective study. Dent Res J (Isfahan) 2014;11:663-675.	Included
Shalish M, Cooper-Kazaz R, Ivgi I, Canetti L, Tsur B, Bachar E, et al. Adult patients' adjustability to orthodontic appliances. Part I: a comparison between Labial, Lingual, and Invisalign. Eur J Orthod 2012;34:724-30.	Included
Soldanova M. [Comparison of treatment effectiveness lingual apparatus 2D and apparatus straight wire]. Master Thesis, 2011, Palacký University, Olomouc, Czech Republic.	Included
Soldanova M, Leseticky O, Komarkova L, Dostalova T, Smutny V, Spidlen M. Effectiveness of treatment of adult patients with the straightwire technique and the lingual two-dimensional appliance. Eur J Orthod 2012;34:674-80.	Included
van der Veen MH, Attin R, Schwestka-Polly R, Wiechmann D. Caries outcomes after orthodontic treatment with fixed appliances: do lingual brackets make a difference? Eur J Oral Sci 2010;118:298-303.	Included
Venkatesh S, Rozario J, Ganeshkar SV, Ajmera S. Comparative evaluation of sagittal anchorage loss in lingual and labial appliances during space closure: A pilot study. APOS Trends Orthod 2015;5:33-7.	Included
Wu A, McGrath C, Wong RW, Wiechmann D, Rabie AB. Comparison of oral impacts experienced by patients treated with labial or customized lingual fixed orthodontic appliances. Am J Orthod Dentofac Orthop 2011;139:784-90.	Included
Wu AK, McGrath C, Wong RW, Wiechmann D, Rabie AB. A comparison of pain experienced by patients treated with labial and lingual orthodontic appliances. Eur J Orthod 2010;32:403-7.	Included

.

Table S3

Additional characteristics of the included trials (supplemental to Table 1)

Trial	Inclusion criteria	Wires for lingual group	Wires for labial group	Sample size justified	
Caniklioglu 2005	NR	0.43mm x 0.43mm CuNiTi	0.43mm x 0.43mm CuNiTi	No sample size calculation	Pat-rep problems: discomfort/ tongue-lip-cheek soreness/ eating, speech, and oral care difficulties/ adaptation period/ general problems (3 mos)
Khattab 2013¥	Class I division 1; 15-30 years; full permanent dentition; moderate upper anterior crowding, indicated for non-extraction treatment; no crossbites; no syndromes; no cleft lip palate; no speech or hearing disorder; no previous orthodontic treatment	0.30mm CuNiTi (individualized; Forestadent®, Germany)	0.30mm CuNiTi (prefabricated; Ormco, Sybron Dental Specialties, Orange, Calif)	Sample size calculated *a priori*	Speech performance evaluated with auditive analysis / Pat-rep oral impairment using a Likert-scale: oral discomfort, speech impairment, mastication difficulties (Bef-Tx, 1 mo, 3 mos)
Khattab 2014¥	Class I division 1; 15-30 years; full permanent dentition; moderate upper anterior crowding, indicated for non-extraction treatment; no crossbites	0.30mm NiTi-0.36mm NiTi-0.41mm CuNiTi (individualized using Template for Biolingual® arches, Forestadent®, Germany)	0.30mm NiTi-0.36mm NiTi-0.41mm CuNiTi (prefabricated Ormco, Sybron Dental Specialties, Orange, Calif)	Sample size calculated (probably *a priori*)	Intercanine-/ interpremolar-/ intermolar width; arch length; amount of enamel reduction (Bef-Tx, after leveling/aligning)
Lombardo 2013	Class I non-extraction patients; 19-23 years; permanent dentition; no caries; no demineralization; no periodontal disease; no antibiotic or antibacterial mouthwash; no systemic disease	NR	NR	No sample size calculation	Oral health parameters: DMFT, PI, GBI / salivary flow rate / salivary buffer capacity and pH / S. mutans count / Lactobacillus count (Bef-Tx, 0 wk, 4 wks, 8 wks)
Rai 2013	Native speakers; 18-35 years; no CLP, no speech or hearing disorders; no previous speech therapy	NR	NR	No sample size calculation	Objective, semiobjective, and subjective speech performance (Bef-Tx, 1 d, 1 wk, 1 mo)
Rai 2014	Native speakers; 18-35 years; moderate crowding; no CLP; no speech or hearing disorders; no previous speech therapy	NR	NR	No sample size calculation	Objective, semiobjective, and subjective speech performance (Bef-Tx, 1 d, 1 wk, 1 mo)
Shalish 2012$	Adult patients; 18-60 years	0.36mm NiTi	0.36mm NiTi	No sample size	Pat-rep health-related quality

				calculation	of life using a Likert-scale / Pat-rep pain intensity and analgesic consumption / number of days needed to achieve mild or no pain (days 1-7 & day 14 after appliance insertion)
Soldanova 2011; 2012	Class I patients with crowding, indicated for non-extraction treatment; completed dental growth	prefabricated 0.30mm NiTi-0.36mm NiTi-0.41mm NiTi-0.41mm SS	Prefabricated 0.30mm NiTi-0.41mm NiTi-0.43mm x 0.51mm NiTi-0.41mm x 0.56mm SS	No sample size calculation	Treatment duration; Intercanine-, interpremolar-, intermolar width / arch length / cephalometric measurements for the sagittal and vertical position of the lower incisors (Bef-Tx, Aft-Tx)
van der Veen 2010	12-18 years; good health; no caries; no demineralizations; fully erupted premolars and canines	NR	NR	Sample size calculated (probably *a priori*)	Number of white spot lesions / (Bef-Tx, After-Tx)
Venkatesh 2015	Bimaxillary dentoalveolar protrusion with planned extraction of upper first premolars; permanent dentition (excluding second-third molars); critical anchorage cases needing 75%-100% anterior retraction; medium-angle cases; no deepbite (no bite ramps should be needed); no severe crowding; no systemic diseases; no syndromes; no Class II or Class III	Prefabricated 0.30mm NiTi-0.36mm NiTi-0.41mm NiTi-0.41mm or 0.46mm SS-0.43mm x 0.51mm TMA	Prefabricated 0.30mm NiTi-0.36mm NiTi-0.41mm NiTi-0.41mm or 0.46mm SS-0.43mm x 0.51mm SS	No sample size calculation	Cephalometric sagittal anchorage loss of the first maxillary permanent molar (before and after space closure)
Wu 2010; 2011	NR	NR	NR	No sample size calculation	Pat-rep pain experience / Pat-rep oral satisfaction: oral discomfort, mastication, speech disturbances, and social functioning using a VAS-scale / Pat-rep sleep disturbance, analgesic consumption, and timing of initial pain (1 wk, 1 mos, 3 mos)

NR, not reported; CuNiTi, copper-nickel-titanium; Pat-rep, patient-reported; mos, months; Bef-Tx, before treatment; DMFT, Decayed, missing, and filled teeth; PI, plaque index; GBI, gingival bleeding index; NiTi, nickel-titanium; SS, stainless steel.wk, week; Aft-Tx, after treatment; VAS, visual analogue scale.

$ a third patient group treated with Invisalign is omitted.

¥ Including also data from communication with the trial's authors.

Table S4

Detailed risk of bias assessment for the included trials

Trial	Sequence generation	Allocation concealment	Blinding of participants, personnel	Blinding of outcome assessors	Incomplete outcome data	Selective outcome reporting	Other sources of bias
Caniklioglu 2005	High risk - no mention of random allocation throughout the paper and highly improbable that it took place: "These patients were divided into two groups".	Unclear - no mention throughout the paper; highly improbable that allocation was concealed.	Unclear - no mention of blinding throughout the paper; blinding is impractical for both patient and clinician.	High risk - no mention of blinding throughout the paper; blinding is practical only for the person responsible for reading the discomfort questionnaires.	Low risk - No drop-outs or patient losses are reported.	Unclear - It is difficult to judge whether selective reporting is a problem, as no protocol exists.	Unclear - residual bias cannot be excluded.
Khattab 2013	Low risk - " patient assignment was based on computer-generated random numbers"	Unclear - allocation concealment probably conducted, but proper description is missing: "allocation procedure was concealed from the researcher and was conducted by one of the co-authors"	Unclear - no mention of blinding throughout the paper; blinding is impractical for both patient and clinician.	High risk - no mention of blinding throughout the paper; blinding should be possible.	Low risk - No drop-outs or patient losses.	Low risk - No protocol exist for the trial. However, given the very specific outcome that the trial was based upon, which is fully reported, it is improbable that selective reporting exists.	Low risk - no additional sources of bias identified.
Khattab 2014	Low risk - "He created a randomization list using Minitab® V.15 with an allocation ratio of 1:1".	Low risk - allocation sequence fully concealed: "The allocation sequence was concealed from the principal researcher (TK) enrolling and assessing participants in sequentially numbered opaque and sealed envelopes. To prevent subversion of the allocation sequence, the name and the date of birth of each participant was written on the envelope and these data were transferred onto the allocation card inside each envelope. Corresponding envelopes were opened only after completing all baseline assessments and the time came to allocate the intervention"	Low risk - blinding is impractical for both patient and clinician; however, authors went to great extent to blind the outcome assessment, which was objective in nature.	Low risk - blinding of outcome assessment adequate: "Blinding of study models to avoid assessor's bias was based on trimming off the brackets from the lingual surfaces of upper teeth on the study models of patients in the lingual group...The opposite procedure was performed for study models belonging to the labial group. Great care was given to make both surfaces of each tooth alike in terms of coarseness. "	Unclear - The authors report that 58 patients were finally considered for randomization, but only 52 patients are analyzed in the end.	Unclear - It is difficult to judge whether selective reporting is a problem, as no protocol could be found.	High risk – bias might have been introduced in the primary outcome (arch dimensions) due to archwire differences (wires were individualized only in the lingual group) and due to differences in the amount of enamel removed.
Lombardo 2013	Unclear - randomization description inadequate: "The 20 non-extraction class I patients were randomly divided into two experimental groups"	Unclear - no mention throughout the paper.	Unclear - no mention of blinding throughout the paper; blinding is impractical for both patient and clinician.	High risk - no mention of blinding throughout the paper; blinding should be possible.	Low risk - No drop-outs or patient losses are reported.	Unclear - It is difficult to judge whether selective reporting is a problem, as no protocol exists.	Unclear - residual bias cannot be excluded.
Rai 2013	High risk - no mention of random allocation throughout the paper and highly improbable that it took place: "The subjects were assigned two different groups — Li (lingual) and La (labial).".	Unclear - no mention throughout the paper.	Unclear - no mention of blinding throughout the paper; blinding is impractical for both patient and clinician.	Unclear – No blinding mentioned, but speech pathologists' assessment (semi-objective) was judged as semi-masked: Two clinical judges trained in speech pathology listened independently to the recording of 59 words, played in a random manner that prevented identification of patients or treatment periods. The same applied for the subjective assessment.	Low risk - No drop-outs or patient losses are reported.	Unclear - It is difficult to judge whether selective reporting is a problem, as no protocol exists.	Unclear - residual bias cannot be excluded.

Rai 2014	High risk - no mention of random allocation throughout the paper and highly improbable that it took place: "The subjects were assigned two different groups — Li (lingual) and La (labial).".	Unclear - no mention throughout the paper.	Unclear - no mention of blinding throughout the paper; blinding is impractical for both patient and clinician.	Low risk – The semi-objective and the subjective speech assessment was reported to be blind. We judged that similar measures were taken also for the objective speech assessment.	Low risk - No drop-outs or patient losses are reported.	Unclear - It is difficult to judge whether selective reporting is a problem, as no protocol exists.	Unclear - residual bias cannot be excluded.
Shalish 2012	High risk - no mention of random allocation throughout the paper and highly improbable that it took place.	Unclear - no mention throughout the paper; highly improbable that allocation was concealed.	Unclear - no mention of blinding throughout the paper; blinding is impractical for both patient and clinician.	High risk - no mention of blinding throughout the paper; blinding is practical only for the person responsible for reading the discomfort questionnaires .	Low risk - No drop-outs or patient losses are reported.	High risk - recovery time not reported in adequate detail. Outcomes from the questionnaire missing from the report.	Unclear - residual bias cannot be excluded.
Soldanova 2011; 2012	High risk - no mention of random allocation throughout the paper and highly improbable that it took place: "These patients were divided into two equal-sized groups according to aesthetic demands of treatment"	Unclear - no mention throughout the paper; highly improbable that allocation was concealed.	Unclear - no mention of blinding throughout the paper; blinding is impractical for both patient and clinician.	High risk - no mention of blinding throughout the paper; blinding could have been implemented.	Low risk - No drop-outs or patient losses are reported.	Unclear - It is difficult to judge whether selective reporting is a problem, as no protocol exists. Included outcomes reported in sufficient detail.	Unclear - residual bias cannot be excluded. Differences in archwire might also introduce bias.
van der Veen 2010	Unclear - randomization description inadequate: "Subjects were randomly appointed to one of two groups"	Unclear - no mention throughout the paper.	Unclear - no mention of blinding throughout the paper; blinding is impractical for both patient and clinician.	High risk - no mention of blinding throughout the paper; blinding could have been implemented.	High risk – Early debracketing of eight patients and two cases of bracket failure, which were not taken into account in the analyses.	High risk - It is difficult to judge whether selective reporting is a problem, as no protocol exists. Included outcome about QLF not reported in sufficient detail.	Unclear - residual bias cannot be excluded.
Venkatesh 2015	High risk - no mention of random allocation throughout the paper and highly improbable that it took place: "The subjects were assigned two different groups — Li (lingual) and La (labial).".	Unclear - no mention throughout the paper.	Unclear - no mention of blinding throughout the paper; blinding is impractical for both patient and clinician.	High risk – No blinding mentioned, although it was feasible.	Low risk - No drop-outs or patient losses are reported.	Unclear - It is difficult to judge whether selective reporting is a problem, as no protocol exists.	Unclear - residual bias cannot be excluded.
Wu 2010; 2011	High risk - no mention of random allocation throughout the paper and highly improbable that it took place.	Unclear - no mention throughout the paper; highly improbable that allocation was concealed.	Unclear - no mention of blinding throughout the paper; blinding is impractical for both patient and clinician.	High risk - no mention of blinding throughout the paper; blinding could have been implemented.	Low risk - No drop-outs or patient losses are reported.	High risk - outcomes not reported in adequate detail.	Unclear - residual bias cannot be excluded.

.

Table S5

Results of individual included trials and performed meta-analyses

Type	Source	Outcome	Studies[†]	Effect	95% CI	P	I^2(%)
Hygiene & soft/hard tissues	Quest	Oral hygiene problems (3 months)	1[a]	RR=1.40	0.91,2.15	0.123	-
	Quest	Oral hygiene problems: high intensity (3 months)	1[a]	RR=0.24	0.01,5.57	0.376	-
	Quest	Oral hygiene problem: food impaction (3 months)	1[a]	RR=1.25	1.03,1.50	**0.022**	-
	Quest	Oral hygiene problem: bleeding gum (3 months)	1[a]	RR=1.38	0.65,2.93	0.410	-
	Quest	Oral hygiene problem: bad taste (3 months)	1[a]	RR=1.40	0.50,3.92	0.522	-
	Clin.	Plaque index (1 month)	1[d]	MD=-0.01	-0.14,0.12	0.879	-
	Clin.	Plaque index (2 months)	1[d]	MD=0.11	0.00,0.22	**0.049**	-
	Clin.	Gingival bleeding index (1 month)	1[d]	MD=-0.10	-0.20,0.00	0.057	-
	Clin.	Gingival bleeding index (2 months)	1[d]	MD=0.09	-0.03,0.21	0.136	-
	Clin./Lab.	Salivary flow rate (1 month)	1[d]	MD=-0.12	-0.45,0.21	0.474	-
	Clin./Lab.	Salivary flow rate (2 months)	1[d]	MD=-0.16	-0.73,0.41	0.579	-
	Clin./Lab.	Salivary buffering capacity (1 month)	1[d]	MD=-0.30	-0.80,0.20	0.235	-
	Clin./Lab.	Salivary buffering capacity (2 months)	1[d]	MD=-0.35	-0.76,0.06	0.097	-
	Clin./Lab.	S. mutans high count (1 month)	1[d]	RR=1.00	0.65,1.55	1.000	-
	Clin./Lab.	S. mutans high count (3 months)	1[d]	RR=1.13	0.78,1.63	0.535	-
	Clin./Lab.	Lactobacillus high count (1 month)	1[d]	RR=2.00	0.68,5.85	0.206	-
	Clin./Lab.	Lactobacillus high count (3 months)	1[d]	RR=1.50	0.60,3.74	0.384	-
	Clin.	Number of new WSL per jaw (univariable*)	1[i]	IRR=0.21	0.08,0.59	**0.003**	-
	Clin.	Number of new WSL per jaw (multivariable*)	1[i]	IRR=0.28	0.10,0.73	**0.010**	-
	Clin.	Sum of change in F per jaw lesions (univariable*)	1[i]	IRR=0.27	0.02,3.22	0.302	-
	Clin.	Sum of change in F per jaw lesions (multivariable*)	1[i]	IRR=0.33	0.03,4.14	0.390	-
	Clin.	Sum of change in Q per jaw lesions (univariable*)	1[i]	IRR=0.07	0.01,0.79	**0.032**	-
	Clin.	Sum of change in Q per jaw lesions (multivariable*)	1[i]	IRR=0.07	0.01,0.69	**0.023**	-
	Clin.	Sum of change in lesion area per jaw lesions (univariable*)	1[i]	IRR=0.06	0.00,0.82	**0.035**	-
	Clin.	Sum of change in lesion area per jaw lesions (multivariable*)	1[i]	IRR=0.13	0.01,1.70	0.121	-
	Clin.	Incidence of new WSL per jaw (univariable*)	1[i]	OR=0.35	0.10,1.22	0.099	-
	Clin.	Incidence of new WSL per jaw (multivariable*)	1[i]	OR=0.43	0.11,1.61	0.209	-
Discomfort	Quest	Oral discomfort (0.5-3 months)	3[abg]	SMD=0.78	0.18,1.38	**0.012**	7
	Quest	Oral discomfort: generalized (3 months)	1[a]	RR=1.45	0.46,4.61	0.529	-
	Quest	Oral discomfort: high intensity (3 months)	1[a]	RR=2.09	0.92,4.76	0.078	-
	Quest	Tongue discomfort (3 months)	1[a]	RR=3.38	1.84,6.18	**<0.001**	-
	Quest	Tongue discomfort: high intensity (3 months)	1[a]	RR=10.61	0.71,159.64	0.088	-
	Quest	Tongue discomfort: duration longer than 30 days (3 months)	1[a]	RR=0.96	0.04,21.65	0.982	-
	Quest	Cheeks discomfort (3 months)	1[a]	RR=0.42	0.26,0.69	**0.001**	-
	Quest	Cheeks discomfort: high intensity (3 months)	1[a]	RR=0.79	0.09,6.77	0.828	-

Quest	Cheeks discomfort: duration longer than 30 days (3 months)	1[a]	RR=0.25	0.02,4.29	0.339	-
Quest	Lips discomfort (3 months)	1[a]	RR=0.36	0.15,0.87	0.023	-
Quest	Lips discomfort: high intensity	1[a]	RR=7.50	0.35,159.54	0.196	-
Quest	Lips discomfort: duration longer than 30 days	1[a]	RR=NE			-
Quest	Oral pain in VAS scale (1 day)	1[g]	MD=0.75	-0.95,2.45	0.387	-
Quest	Oral pain in VAS scale (2 days)	1[g]	MD=1.86	0.02,3.71	**0.048**	-
Quest	Oral pain in VAS scale (3 days)	1[g]	MD=1.86	0.04,3.68	**0.045**	-
Quest	Oral pain in VAS scale (4 days)	1[g]	MD=1.78	0.11,3.45	**0.036**	-
Quest	Oral pain in VAS scale (5 days)	1[g]	MD=1.61	0.12,3.10	**0.034**	-
Quest	Oral pain in VAS scale (6 days)	1[g]	MD=2.11	0.64,3.58	**0.005**	-
Quest	Oral pain in VAS scale (7 days)	1[g]	MD=1.61	0.37,2.85	**0.011**	-
Quest	Oral pain in VAS scale (14 days)	1[g]	MD=1.19	0.27,2.11	**0.012**	-
Quest	Oral discomfort in Likert scale (1 day)	1[g]	MD=0.49	-0.09,1.07	0.098	-
Quest	Oral discomfort in Likert scale (2 days)	1[g]	MD=0.86	0.22,1.50	**0.008**	-
Quest	Oral discomfort in Likert scale (3 days)	1[g]	MD=0.92	0.29,1.55	**0.004**	-
Quest	Oral discomfort in Likert scale (4 days)	1[g]	MD=0.82	0.27,1.37	**0.003**	-
Quest	Oral discomfort in Likert scale (5 days)	1[g]	MD=0.69	0.19,1.19	**0.007**	-
Quest	Oral discomfort in Likert scale (6 days)	1[g]	MD=0.84	0.34,1.34	**0.001**	-
Quest	Oral discomfort in Likert scale (7 days)	1[g]	MD=0.67	0.26,1.08	**0.001**	-
Quest	Oral discomfort in Likert scale (14 days)	1[g]	MD=0.42	-0.03,0.87	0.068	-
Quest	Appliance adaptation: duration longer than 30 days (3 months)	1[a]	RR=7.00	0.38,129.93	0.192	-
Quest	Appliance handling: very difficult (3 months)	1[a]	RR=1.00	0.36,2.75	1.000	-
Quest	Appliance noticibility (3 months)	1[a]	RR=0.31	0.18,0.53	**<0.001**	-
Quest	Irritation of soft tissues (0 mo)	1[b]	RR=1.84	1.18,2.87	**0.007**	-
Quest	Irritation of soft tissues (1 month)	1[b]	RR=1.75	0.63,4.89	0.286	-
Quest	Irritation of soft tissues (3 months)	1[b]	RR=7.00	0.39,125.99	0.187	-
Quest	General activity problems in Likert scale (1 day)	1[g]	MD=0.49	-0.02,1.00	0.058	-
Quest	General activity problems in Likert scale (2 days)	1[g]	MD=0.74	0.16,1.32	**0.013**	-
Quest	General activity problems in Likert scale (3 days)	1[g]	MD=0.65	0.13,1.17	**0.014**	-
Quest	General activity problems in Likert scale (4 days)	1[g]	MD=0.70	0.22,1.18	**0.004**	-
Quest	General activity problems in Likert scale (5 days)	1[g]	MD=0.47	0.04,0.90	**0.033**	-
Quest	General activity problems in Likert scale (6 days)	1[g]	MD=0.48	0.15,0.81	**0.004**	-
Quest	General activity problems in Likert scale (7 days)	1[g]	MD=0.47	0.12,0.82	**0.009**	-
Quest	General activity problems in Likert scale (14 days)	1[g]	MD=0.25	-0.05,0.55	0.097	-
Quest	Oral symptoms in Likert scale (1 day)	1[g]	MD=-0.04	-0.44,0.36	0.845	-
Quest	Oral symptoms in Likert scale (2 days)	1[g]	MD=0.22	-0.16,0.60	0.252	-
Quest	Oral symptoms in Likert scale (3 days)	1[g]	MD=0.17	-0.22,0.56	0.393	-
Quest	Oral symptoms in Likert scale (4 days)	1[g]	MD=0.15	-0.20,0.50	0.400	-
Quest	Oral symptoms in Likert scale (5 days)	1[g]	MD=0.13	-0.24,0.50	0.494	-
Quest	Oral symptoms in Likert scale (6 days)	1[g]	MD=0.15	-0.20,0.50	0.405	-
Quest	Oral symptoms in Likert scale (7 days)	1[g]	MD=0.25	-0.08,0.58	0.136	-

	Quest	Oral symptoms in Likert scale (14 days)	1[g]	MD=0.02	-0.31,0.35	0.905	-
	Quest	Sleep disturbance (1 week)	1[k]	RR=1.00	0.15,6.64	1.000	-
	Quest	Sleep disturbance (1 month)	1[k]	RR=1.50	0.27,8.34	0.643	-
	Quest	Sleep disturbance (3 months)	1[k]	RR=0.33	0.01,7.87	0.496	-
	Quest	Analgesic consumption (1 week)	1[k]	RR=4.00	0.47,33.73	0.203	-
	Quest	Analgesic consumption (1 month)	1[k]	RR=5.00	0.25,99.95	0.292	-
	Quest	Analgesic consumption (3 months)	1[k]	RR=NE	-	-	-
	Quest	Pain pattern: mainly during day (1 week)	1[k]	RR=1.00	0.46,2.17	1.000	-
	Quest	Pain pattern: mainly during day (1 month)	1[k]	RR=0.80	0.24,2.69	0.719	-
	Quest	Pain pattern: mainly during day (3 months)	1[k]	RR=1.00	0.22,4.56	1.000	-
	Quest	Pain pattern: same day and night (1 week)	1[k]	RR=1.00	0.60,1.66	1.000	-
	Quest	Pain pattern: same day and night (1 month)	1[k]	RR=1.33	0.95,1.88	0.100	-
	Quest	Pain pattern: same day and night (3 months)	1[k]	RR=1.00	0.80,1.25	1.000	-
	Quest	Initial pain: after 3 hours (1 week)	1[k]	RR=0.04	0.01,0.25	**0.001**	-
	Quest	Initial pain: after 3 hours (1 month)	1[k]	RR=0.05	0.01,0.23	**<0.001**	-
	Quest	Initial pain: after 3 hours (3 months)	1[k]	RR=0.15	0.06,0.35	**<0.001**	-
Speech	Aud. Anal.	Upper boundary frequency of the /s/ sound in the middle of the word (0 mo)	3[bef]	MD=-722.29	-1500.00,94.26	0.083	97
	Aud. Anal.	Upper boundary frequency of the /s/ sound in the middle of the word (1 month)	3[bef]	MD=-441.12	-986.22,103.98	0.113	95
	Aud. Anal.	Upper boundary frequency of the /s/ sound in the middle of the word (1 week)	2[ef]	MD=-312.18	-600.98,-23.39	**0.034**	75
	Aud. Anal.	Upper boundary frequency of the /s/ sound in the middle of the word (3 months)	1[b]	MD=-443.00	-647.06,-238.94	**<0.001**	-
	Aud. Anal.	Upper boundary frequency of the /s/ sound in the start of the word (1 day)	2[ef]	MD=-39.87	-195.65,115.91	0.616	0
	Aud. Anal.	Upper boundary frequency of the /s/ sound in the start of the word (1 week)	2[ef]	MD=-216.93	-372.62,-61.25	**0.006**	0
	Aud. Anal.	Upper boundary frequency of the /s/ sound in the start of the word (1 month)	2[ef]	MD=-120.22	-282.73,42.29	0.147	0
	Clin.	Speech performance by expert (1 day)	1[e]	MD=0.02	-0.04,0.08	0.497	-
	Clin.	Speech performance by expert (1 week)	1[e]	MD=0.24	0.13,0.35	**<0.001**	-
	Clin.	Speech performance by expert (1 month)	1[e]	MD=0.07	0.01,0.13	**0.019**	-
	Clin.	Speech performance by layperson (1 day)	2[ef]	MD=0.00	-0.22,0.21	0.989	-
	Clin.	Speech performance by layperson (1 week)	2[ef]	MD=0.82	0.58,1.05	**<0.001**	-
	Clin.	Speech performance by layperson (1 month)	2[ef]	MD=0.60	0.31,0.90	**<0.001**	-
	Clin.	Speech pathologist's assessment of vowels (1 day)	1[f]	MD=-0.25	-1.73,1.23	0.741	-
	Clin.	Speech pathologist's assessment of vowels (1 week)	1[f]	MD=2.88	0.58,5.18	**0.014**	-
	Clin.	Speech pathologist's assessment of vowels (1 month)	1[f]	MD=0.12	-0.58,0.82	0.737	-
	Clin.	Speech pathologist's assessment of palatal consonants (1 day)	1[f]	MD=1.02	-0.58,2.62	0.210	-
	Clin.	Speech pathologist's assessment of palatal consonants (1 week)	1[f]	MD=2.19	0.93,3.45	**0.001**	-
	Clin.	Speech pathologist's assessment of palatal consonants (1 month)	1[f]	MD=0.29	0.13,0.45	**<0.001**	-
	Clin.	Speech pathologist's assessment of alveolar consonants (1 day)	1[f]	MD=1.79	0.55,3.03	**0.005**	-
	Clin.	Speech pathologist's assessment of alveolar consonants (1 week)	1[f]	MD=2.81	1.56,4.06	**<0.001**	-
	Clin.	Speech pathologist's assessment of alveolar consonants (1 month)	1[f]	MD=0.35	-0.05,0.75	0.087	-
	Clin.	Speech pathologist's assessment of blends (1 day)	1[f]	MD=0.25	-0.28,0.78	0.351	-

	Clin.	Speech pathologist's assessment of blends (1 week)	1[f]	MD=0.83	0.16,1.50	**0.015**	-
	Clin.	Speech pathologist's assessment of blends (1 month)	1[f]	MD=0.04	-0.02,0.10	0.215	-
	Quest	Speech disturbance (3 months)	2[ab]	RR=8.90	1.15,68.69	**0.036**	0
	Quest	Speech disturbance: high intensity (3 months)	1[a]	RR=10.97	0.67,179.68	0.093	-
	Quest	Speech disturbance: duration longer than 30 days (3 months)	1[a]	RR=9.68	0.58,160.26	0.113	-
	Quest	Perception of articulation change (0 month)	1[b]	RR=2.33	1.34,4.05	**0.003**	-
	Quest	Perception of articulation change (1 month)	1[b]	RR=25.00	1.60,391.13	**0.022**	-
	Quest	Perception of articulation change (3 months)	1[b]	RR=11.00	0.66,184.62	0.096	-
	Quest	Others' observation of articulation change (0 month)	1[b]	RR=4.00	1.68,9.50	**0.002**	-
	Quest	Others' observation of articulation change (1 month)	1[b]	RR=15.00	0.92,243.52	0.057	-
	Quest	Others' observation of articulation change (3 months)	1[b]	RR=5.00	0.26,97.00	0.287	-
	Quest	Avoidance of some types of conversations (0 month)	1[b]	RR=4.50	1.14,17.83	**0.032**	-
	Quest	Avoidance of some types of conversations (1 month)	1[b]	RR=7.00	0.39,125.99	0.187	-
	Quest	Avoidance of some types of conversations (3 months)	1[b]	RR=NE	-	-	-
Eating	Quest	Eating problems (3 months)	2[ab]	RR=5.35	0.97,29.50	0.054	0
	Quest	Eating problems: high intensity (3 months)	1[a]	RR=1.35	0.69,2.63	0.385	-
	Quest	Eating problems: duration longer than 30 days (3 months)	1[a]	RR=3.72	0.44,31.27	0.226	-
	Quest	Eating hard foods problems (3 months)	1[a]	RR=1.17	0.95,1.43	0.136	-
	Quest	Eating hard foods problems: high intensity (3 months)	1[a]	RR=1.29	0.63,2.61	0.487	-
	Quest	Eating soft foods problems (3 months)	1[a]	RR=2.00	0.94,4.25	0.071	-
	Quest	Eating soft foods problems: high intensity (3 months)	1[a]	RR=NE			-
	Quest	Mastication problems (0 month)	1[b]	RR=1.30	0.98,1.71	0.067	-
	Quest	Mastication problems (1 month)	1[b]	RR=2.25	0.86,5.92	0.100	-
	Quest	Mastication problems (3 months)	1[b]	RR=9.00	0.52,155.24	0.130	-
	Quest	Eating problems in Likert scale (1 day)	1[g]	MD=0.74	0.18,1.30	**0.010**	-
	Quest	Eating problems in Likert scale (2 days)	1[g]	MD=1.05	0.45,1.65	**0.001**	-
	Quest	Eating problems in Likert scale (3 days)	1[g]	MD=1.09	0.46,1.72	**0.001**	-
	Quest	Eating problems in Likert scale (4 days)	1[g]	MD=1.08	0.44,1.72	**0.001**	-
	Quest	Eating problems in Likert scale (5 days)	1[g]	MD=1.12	0.48,1.76	**0.001**	-
	Quest	Eating problems in Likert scale (6 days)	1[g]	MD=1.21	0.61,1.82	**<0.001**	-
	Quest	Eating problems in Likert scale (7 days)	1[g]	MD=1.11	0.50,1.72	**<0.001**	-
	Quest	Eating problems in Likert scale (14 days)	1[g]	MD=1.11	0.49,1.73	**<0.001**	-
Dental form	Model	Intercanine width (after Tx)	2[ch]	MD=0.62	0.08,1.16	**0.025**	0
	Model	Interpremolar width (after Tx)	2[ch]	MD=-1.47	-3.41,0.48	0.139	87
	Model	Intermolar width (after Tx)	2[ch]	MD=-0.63	-2.45,1.19	0.499	86
	Model	Arch length to molars (after Tx)	2[ch]	MD=-0.48	-1.75,0.79	0.461	70
	Model	Amount of Enamel Reduction (after Tx)	1[c]	MD=-0.67	-0.84,-0.50	**<0.001**	-
	Model	Arch length to canines (after Tx)	1[h]	MD=-0.26	-0.91,0.39	0.430	-

	Model	Arch length: lateral segment (after Tx)	1[h]	MD=0.54	-0.61,1.69	0.358	-
Cephalometric	Ceph.	Sagittal position of the lower incisor to the A-Pg line (after Tx)	1[h]	MD=1.04	-0.00,2.08	0.051	-
	Ceph.	Sagittal position of the lower incisor apex on the mandibular plane (after Tx)	1[h]	MD=-1.16	-4.41,2.09	0.484	-
	Ceph.	Inclination of the lower incisor to the mandibular plane (after Tx)	1[h]	MD=0.08	-1.34,1.50	0.912	-
	Ceph.	Y axix (sella turcica to gnathion) (after Tx)	1[h]	MD=0.68	-0.57,1.93	0.287	-
	Ceph.	Lower anterior face height (after Tx)	1[h]	MD=0.20	-0.94,1.34	0.730	-
	Ceph.	Upper lip to esthetic line (after Tx)	1[h]	MD=0.00	-0.74,0.74	1.000	-
	Ceph.	Lower lip to esthetic line (after Tx)	1[h]	MD=0.60	-0.49,1.69	0.280	-
	Ceph.	Sagittal anchorage loss of the upper first molar (after Tx)	1[j]	MD=-0.82	-1.09,-0.56	**<0.001**	-
Excluded	Quest	Oral hygiene problems (Likert scale)	0[g]	Missing data	-	-	-
	Quest	Sever oral pain	0[g]	Missing data	-	-	-
	Quest	Recovery time for each problem	0[g]	Missing data	-	-	-
	Quest	Oral pain at tongue (VAS scale)	0[k]	Missing data	-	-	-
	Quest	Oral pain at cheeks (VAS scale)	0[k]	Missing data	-	-	-
	Quest	Oral pain at lips (VAS scale)	0[k]	Missing data	-	-	-
	Quest	Oral pain at gums (VAS scale)	0[k]	Missing data	-	-	-
	Quest	Oral pain at face (VAS scale)	0[k]	Missing data	-	-	-
	Quest	Oral pain at jaw (VAS scale)	0[k]	Missing data	-	-	-
	Quest	Total oral pain during the treatment (VAS scale)	0[k]	Missing data	-	-	-
	Quest	Sleep disturbance	0[k]	Missing data	-	-	-
	Quest	Analgesic consumption	0[k]	Missing data	-	-	-
	Quest	Pain pattern: day vs night vs both	0[k]	Missing data	-	-	-
	Quest	Timing of first pain: <3 hours vs > 3hours	0[k]	Missing data	-	-	-

RR, relative risk; MD, mean difference; IRR, incidence rate ratio; OR, odds ratio; SMD, standardized mean difference; CI, confidence interval; NE, not estimable; WSL, white spot lesion; F, average fluorescence loss within a lesion relative to the fluorescence level of healthy tissue surrounding the lesion; Q, integrated fluorescence loss over the lesion; VAS, visual analogue scale; Tx, treatment

[†]a, Caniklioglu 2005; b, Khattab 2013; c, Khattab 2014; d, Lombardo 2013; e, Rai 2013; f, Rai 2014; g, Shalish 2012; h, Soldanova 2011 and 2012; i, van der Veen 2010; j, Venkatesh 2015; k, Wu 2010 and 2011.

*Generalized estimating equations adopting a negative binomial distribution for continuous outcomes and binomial distribution for binary outcomes were used to calculate newly formed white spot lesions on the treated side of each jaw, while accounting for within-patient clustering. Multivariable estimates correspond to IRRs or ORs adjusted for type of jaw (maxilla or mandible), patient age, treatment duration, and number of baseline lesions.

Table S6

Details of the GRADE assessment for the main outcomes of this systematic review (outcome numbering corresponds to the order of Table 3)

Outcome	Risk of Bias	Inconsistency	Indirect-ness	Imprecision	Publication bias	Large Effect	Dose Response	Residual Confounding
Outcome 1	*Starts from "low", due to the inclusion of non-randomized studies. Downgraded further by one point due to serious limitations (high risk of bias).*	*High heterogeneity; confidence regarding decision unaffected; heterogeneity affects just the precision of the estimate.*	*Directly relevant*	*Adequate sample*	*No evidence of bias*	*No reason to rate up*	*No dose response relation assessment.*	*Cannot be ruled out.*
Outcome 2	*Same as Outcome 1*	*High heterogeneity; confidence regarding decision unaffected; heterogeneity affects just the precision of the estimate.*	*Same as Outcome 1*	*Inadequate sample; the 95% CI includes both the null effect and large effect values, which indicates imprecision.*	*Same as Outcome 1*	*Large effect magnitude; however no rating up due existing concerns regarding risk of bias and imprecision*	*Same as Outcome 1*	*Same as Outcome 1*
Outcome 3	*Same as Outcome 1*	*Low heterogeneity; no reason to downgrade*	*Same as Outcome 1*	*Adequate sample*	*Same as Outcome 1*	*No reason to rate up*	*Same as Outcome 1*	*Same as Outcome 1*
Outcome 4	*Same as Outcome 1*	*Low heterogeneity; no reason to downgrade*	*Same as Outcome 1*	*Same as Outcome 1*	*Same as Outcome 1*	*Same as Outcome 2*	*Same as Outcome 1*	*Same as Outcome 1*
Outcome 5	*Same as Outcome 1*	*Low heterogeneity; no reason to downgrade*	*Outcome not necessarily directly relevant. Treatment mechanics and wire might have confounded the results.*	Adequate sample	*Same as Outcome 1*	*No reason to rate up*	*Same as Outcome 1*	*Same as Outcome 1*
Outcome 6	*Same as Outcome 1*	*High heterogeneity, which could not be explained by subgroup analysis, while our confidence regarding decision is affected by it (trials on both sides of the forest plot)*	*Same as Outcome 5*	*Adequate sample*	*Same as Outcome 1*	*No reason to rate up*	*Same as Outcome 1*	*Same as Outcome 1*
Outcome 7	*Same as Outcome 1*	*No heterogeneity assessment*	*Same as Outcome 1*	*Adequate sample*	*Same as Outcome 1*	*No reason to rate up*	*Same as Outcome 1*	*Same as Outcome 1*

.

Table S7

GRADE summary of findings table for the main outcomes of the systematic review according to the sensitivity analysis

	Sensitivity analysis						Original analysis	
	Illustrative comparative effects (95% CI)							
Outcomes	**Labial appliances**	**Lingual appliances**	**Change**	**Patients (trials)**	**GRADE**	**Effects**	**GRADE**	**Effects**
Oral discomfort; patient-reported (0.5-3 months)	Assumed risk 529 patient per 1000	Corresponding risk 1648 more patients per 1000 (204 to 5937 more)	2 pCCT omitted	34 (1)	Moderate ⊕⊕⊕⊖[1]	SMD=1.90 (0.27,3.53); P<0.05	Very low ⊖⊖⊖⊖	SMD=0.78 (0.18,1.38); P<0.05
Upper boundary frequency of the /s/ sound in the middle of the word; auditory analysis (1 month)	Assumed change The upper boundary frequency decreased on average by 114.80 Hz in the labial groups	Corresponding change Decrease by 1096.20 Hz (95% CI: 830.45 Hz to 1361.95 Hz decrease) compared to the labial group.	2 pCCTs omitted	34 (1)	Moderate ⊕⊕⊕⊖[2]	MD=-1096.20 (-1361.95,-830.45); P<0.05	Very low ⊖⊖⊖⊖	MD=-441.12 (-986.22,203.98) ; P>0.05
Speech performance; assessed by layperson (1 month)	Assumed change -	Corresponding change -	2 pCCTs omitted	0 (0)	-	-	Very low ⊖⊖⊖⊖	MD=0.60 (0.31,0.90); P<0.05
Eating difficulty; patient-reported (3 months)	Assumed risk 31 patients per 1000*	Corresponding risk 264 more patients per 1000 (16 less to 5090 more)	1 pCCT omitted	34 (1)	Moderate ⊕⊕⊕⊖[2]	RR=9.00 (0.52,155.24); P>0.05	Very low ⊖⊖⊖⊖	RR=5.35 (0.97,29.50); P>0.05
Intercanine width; from dental cast analysis (after treatment)	Assumed change Increase on average by 1.30 mm in the labial group	Corresponding change Increase by 0.69 mm (95% CI: 0.03 mm to 1.35 mm increase) compared to the labial group.	1 pCCT omitted	52 (1)	Moderate ⊕⊕⊕⊖[3]	MD=0.69 (0.03,1.35); P<0.05	Very low ⊖⊖⊖⊖	MD=0.62 (0.08,1.16); P<0.05
Intermolar width; from dental cast analysis (after treatment)	Assumed change Increase on average by 0.80 mm in the labial group	Corresponding change Decrease by 1.59 mm (95% CI: 0.52 mm to 2.66 mm decrease) compared to the labial group.	1 pCCT omitted	52 (1)	Moderate ⊕⊕⊕⊖[2]	MD=-1.59 (-2.66,-0.52); P<0.05	Very low ⊖⊖⊖⊖	MD=-0.63 (-2.45,1.19); P>0.05
Sagittal anchorage loss of the upper first molar during space closure (after treatment)	Assumed change -	Corresponding change -	1 pCCT omitted	0 (0)	-	-	Very low ⊖⊖⊖⊖	MD=-0.82 (-1.09,-0.56); P<0.05

CI, confidence interval; MD, mean differences; SMD, standardized mean differences; pCCT, prospective non-randomized clinical trial; RR, relative risk; Tx, treatment.

*assumed risk extracted from the omitted trial, as the included trial had no events in the control group

[1]GRADE starts from high as only randomized trials are included; downgraded by one due to absence of blinding.

[2]GRADE starts from high as only randomized trials are included; downgraded by one due to absence of blinding; very large effect magnitude, but no upgrade, due to existing limitations.

[3]GRADE starts from high as only randomized trials are included; downgraded by one due to the possibility of residual confounding (archwire differences between the lingual and the labial groups might have influenced the dental arch form and width).

Table S8

Supplementary Information

Communications with trialists

- Dr. Shalish (trial Shalish 2011) was contacted for clarifications, any extra trial data, and missed trials: responded that trial was expanded for publication of additional papers and no data could be provided.
- Dr. Rai (trials Rai 2013; Rai 2014) contacted for clarifications, any extra trial data, and missed trials: responded that he would be willing to help and received modified extraction form with questions; waiting for response.
- Dr. Khattab (trials Khattab 2013; Khattab 2014) contacted for clarifications, any extra trial data, and missed trials: responded that he would be willing to help and received modified extraction form with questions; provided clarifications about wire/bracket information and confirmed that the two trials are unique.
- Dr. Soldanova (trial Soldanova 2011; 2012) contacted for clarifications, any extra trial data, and missed trials: responded that she would be willing to help and received modified extraction form with questions; waiting for response.
- Dr. Caniklioglu (trial Caniklioglu 2013) contacted for clarifications, any extra trial data, and missed trials: did not respond to e-mail. Contacted the second author, Dr. Oztürk with the same query: responded that trial is of prospective nature and can be included.
- Dr. van der Veen (trial van der Veen 2010) contacted for raw trial data: responded by sending all raw data in SPSS file format.
- Dr. Rabie (trials Wu 2010, 2011) contacted for raw trial data, clarifications, any extra trial data, and missed trials: Dr. Rabie responded that he has retired. Prompted me to contact Dr. McGrath. Dr. McGrath contacted, but did not respond.

Author contributions

SNP conceived the idea and wrote the first draft of the protocol. SNP, LG, AJ, TE, and CB revised the protocol. SNP performed the literature searches, extracted search hits, and did screening by title. SNP and LG did study selection by abstract and full-text, did data extraction, and assessed the risk of bias in duplicate, while AJ, TE, and CB resolved any conflicts that arose. SNP handled communications with trialists, performed the statistical analysis, and wrote the first draft of the manuscript. SNP, LG, AJ, TE, and CB assisted in the interpretation of the results and revised the manuscript draft. SNP submitted the manuscript, is the guarantor and responsible for the accuracy of the data and for future updates of the review.

***Post hoc* changes to the protocol**

- The outcomes chosen for the GRADE analyses were modified, according to the trials that were identified.
- Contrary to the original analysis plan, we used odds ratios and incidence rate ratios for one identified trial for the analysis, as the raw data from this split-mouth trial were re-analyzed through multivariable regression modeling adopting a binomial or negative binomial distribution, as appropriately.
- We used the standardized mean difference as effect measure in one instance, as binary and continuous measurements of oral discomfort were combined into the same meta-analysis.
- Subgroup analyses and assessments of reporting biases were planned, but could not be performed due to the limited number of trials included in the meta-analyses. We could not perform any subgroup analyses, as less than 5 trials were included in every meta-analysis, and trial characteristics overlapped with study design characteristics.
- The number needed to treat was planned to be used to clinically translate the results of statistically significant meta-analyses of binary outcomes, but no significant binary outcomes existed.

- We added an extra exploratory analysis based on meta-epidemiological methods to further the planned sensitivity analysis of including only RCT. This is clearly denoted as an exploratory analysis in the sensitivity analysis.